No easy answer

Demands to analyse Connecticut school shooter's DNA are misguided and could lead to dangerous stigmatization, or worse.

Connecticut's state medical examiner has requested a full genetic analysis of mass killer Adam Lanza, who shot 20 children, 6 school staff, his mother and himself in Newtown in December. At first glance, it is easy to understand why. Confronted with such senseless violence, it is human nature to seek solace in scientific explanations. After John Wayne Gacy was executed in 1994 for the murder of 33 young men and boys, his brain was preserved and examined for clues to what made him a monster. More than 80 years ago, scientists reportedly studied the brain of serial killer Peter Kürten, the 'vampire of Dusseldorf', who was executed in 1931.

This quest to understand endures as technology advances. Now, instead of looking at cranial folds and frontal lobes for clues to the massacre, geneticists at the University of Connecticut in Farmington will scour Lanza's genes. On its own, this hunt will be about as informative as studies of the brains of murderers: not very.

The Connecticut scientists will not talk about the job they have been handed. It is not clear what they will find, or even what they should look for. Suspend disbelief for a moment and pretend that a 'mass-shooter gene' exists — something that no serious geneticist believes — and scientists could still draw no conclusions from a single individual's genome.

To be sure, many links and suggestions of links have been identified between genetics, mental illness and, to a lesser extent, violence. A study using Swedish registries (R. Kuja-Halkola *et al. Dev. Psychopathol.* **24**, 739–753; 2012) found that children born to men older than 60 were more likely to be convicted of violent crimes than were those born to men aged 40–60 years, an observation that might be linked to increasing numbers of mutations in sperm as men age. Genetic risk factors have been identified for autism, depression and schizoid spectrum disorders, but they explain relatively little. People diagnosed with schizoid spectrum disorders are more likely to be convicted of violent crimes than are those with no such diagnosis, but the vast majority of people with mental illness do not commit crimes.

Such associations hold only for groups. Many healthy people carry mutations associated with disease; many people with mental illness carry no known risk factors. Mass shooters are often young white men, yet very few young white men become mass shooters. There is no one-to-one relationship between genetics and mental health or between mental health and violence. Something as simple as a DNA sequence cannot explain anything as complex as behaviour.

But there is a dangerous tendency to oversimplify, especially in the wake of tragedy. If Lanza's DNA reveals genetic variants — as it inevitably will — people who carry similar variants could be stigmatized, even if those variants are associated only with ear shape. If Lanza has genetic variants already associated with autism or depression, people with those diseases could come under suspicion as well.

The real risk here, and the real flaw in the Connecticut exercise, is that to identify a genetic variant is more straightforward — but arguably less informative — than to characterize the complex environment of the individual. Lanza's DNA will be analysed not because it will be useful but because it can be analysed. The ease of DNA sequencing will lead to a dangerous temptation to focus on minor, even spurious, genetic correlations at the expense of non-genetic factors that are more influential.

Geneticists must explain — and in the wake of the Lanza move many already are — that the ability to sequence DNA is many steps removed from the ability to make that sequence meaningful. Many, if not most, mutations are meaningless outside narrow contexts. One of the most robust examples of a variant that has been linked to antisocial behaviour holds only for individuals who experience severe childhood trauma or abuse; those who do not face no greater risk of being antisocial than people without the variant. Less-studied risk factors presumably work the same way. Genetics matters only in the context of environment.

"This hunt will be about as informative as studies of the brains of murderers: not very."

Research must go beyond the drive to unpick impulses to violence. It should consider the means to violence. On the day that Lanza entered Sandy Hook school armed with his mother's guns, another deranged man entered Chenpeng Village school in China's Henan province armed with a knife. Horrific as that attack was, no one was killed. Yet research on how to reduce gun violence in the United States has been hampered by pro-gun lobbyists and politicians, who have reduced the necessary funding and record-keeping for preventing injuries from firearms.

Sequencing Lanza's DNA may bring benefits, but only if appropriate expectations are maintained. At best, such studies will prompt deeper interest and funding to figure out the biological and social factors that really harm. At worst, they will stigmatize those with mental illnesses and distract scientists from more-important questions. ■

Culture shock

Health-benefit claims for Europe's foods must at last be substantiated by science.

Will a daily probiotic yogurt improve your immune defences? And will cooking with olive oil boost the levels of 'good' cholesterol in your blood?

So far, the food companies behind these particular claims have not supplied the hard scientific evidence to convince expert committees of the European Food Safety Authority (EFSA) in Parma, Italy, that they are warranted. European Union (EU) legislation means that all similarly unsubstantiated health claims for food will soon be disallowed. A register of permissible claims — which will be regularly updated

THIS WEEK

EDITORIALS

X FACTORS The global risks we know we don't know about **p.134**

WORLD VIEW Rumour psychology and a tale of two airlines **p.135**

BEHAVIOUR Female Atlantic mollies go for bisexual males **p.136**

No easy answer

Demands to analyse Connecticut school shooter's DNA are misguided and could lead to dangerous stigmatization, or worse.

Connecticut's state medical examiner has requested a full genetic analysis of mass killer Adam Lanza, who shot 20 children, 6 school staff, his mother and himself in Newtown in December. At first glance, it is easy to understand why. Confronted with such senseless violence, it is human nature to seek solace in scientific explanations. After John Wayne Gacy was executed in 1994 for the murder of 33 young men and boys, his brain was preserved and examined for clues to what made him a monster. More than 80 years ago, scientists reportedly studied the brain of serial killer Peter Kürten, the 'vampire of Dusseldorf', who was executed in 1931.

This quest to understand endures as technology advances. Now, instead of looking at cranial folds and frontal lobes for clues to the massacre, geneticists at the University of Connecticut in Farmington will scour Lanza's genes. On its own, this hunt will be about as informative as studies of the brains of murderers: not very.

The Connecticut scientists will not talk about the job they have been handed. It is not clear what they will find, or even what they should look for. Suspend disbelief for a moment and pretend that a 'mass-shooter gene' exists — something that no serious geneticist believes — and scientists could still draw no conclusions from a single individual's genome.

To be sure, many links and suggestions of links have been identified between genetics, mental illness and, to a lesser extent, violence. A study using Swedish registries (R. Kuja-Halkola *et al. Dev. Psychopathol.* **24**, 739–753; 2012) found that children born to men older than 60 were more likely to be convicted of violent crimes than were those born to men aged 40–60 years, an observation that might be linked to increasing numbers of mutations in sperm as men age. Genetic risk factors have been identified for autism, depression and schizoid spectrum disorders, but they explain relatively little. People diagnosed with schizoid spectrum disorders are more likely to be convicted of violent crimes than are those with no such diagnosis, but the vast majority of people with mental illness do not commit crimes.

Such associations hold only for groups. Many healthy people carry mutations associated with disease; many people with mental illness carry no known risk factors. Mass shooters are often young white men, yet very few young white men become mass shooters. There is no one-to-one relationship between genetics and mental health or between mental health and violence. Something as simple as a DNA sequence cannot explain anything as complex as behaviour.

But there is a dangerous tendency to oversimplify, especially in the wake of tragedy. If Lanza's DNA reveals genetic variants — as it inevitably will — people who carry similar variants could be stigmatized, even if those variants are associated only with ear shape. If Lanza has genetic variants already associated with autism or depression, people with those diseases could come under suspicion as well.

The real risk here, and the real flaw in the Connecticut exercise, is that to identify a genetic variant is more straightforward — but arguably less informative — than to characterize the complex environment of the individual. Lanza's DNA will be analysed not because it will be useful but because it can be analysed. The ease of DNA sequencing will lead to a dangerous temptation to focus on minor, even spurious, genetic correlations at the expense of non-genetic factors that are more influential.

Geneticists must explain — and in the wake of the Lanza move many already are — that the ability to sequence DNA is many steps removed from the ability to make that sequence meaningful. Many, if not most, mutations are meaningless outside narrow contexts. One of the most robust examples of a variant that has been linked to antisocial behaviour holds only for individuals who experience severe childhood trauma or abuse; those who do not face no greater risk of being antisocial than people without the variant. Less-studied risk factors presumably work the same way. Genetics matters only in the context of environment.

> ***"This hunt will be about as informative as studies of the brains of murderers: not very."***

Research must go beyond the drive to unpick impulses to violence. It should consider the means to violence. On the day that Lanza entered Sandy Hook school armed with his mother's guns, another deranged man entered Chenpeng Village school in China's Henan province armed with a knife. Horrific as that attack was, no one was killed. Yet research on how to reduce gun violence in the United States has been hampered by pro-gun lobbyists and politicians, who have reduced the necessary funding and record-keeping for preventing injuries from firearms.

Sequencing Lanza's DNA may bring benefits, but only if appropriate expectations are maintained. At best, such studies will prompt deeper interest and funding to figure out the biological and social factors that really harm. At worst, they will stigmatize those with mental illnesses and distract scientists from more-important questions. ■

Culture shock

Health-benefit claims for Europe's foods must at last be substantiated by science.

Will a daily probiotic yogurt improve your immune defences? And will cooking with olive oil boost the levels of 'good' cholesterol in your blood?

So far, the food companies behind these particular claims have not supplied the hard scientific evidence to convince expert committees of the European Food Safety Authority (EFSA) in Parma, Italy, that they are warranted. European Union (EU) legislation means that all similarly unsubstantiated health claims for food will soon be disallowed. A register of permissible claims — which will be regularly updated

whenever new scientific evidence can be brought to bear — will be presented to the European Parliament for inspection next month, and adopted by member states later this year. Already, many health-food companies, wary of possible litigation, have toned down their marketing claims.

The regulations, drafted by the European Commission and adopted by the EU in 2007, are intended to put an end to the free ride that the food and food-supplement industries have enjoyed until now. Companies will no longer be able to market their products with unproven promises that consumers will be healthier, slimmer or have happier lives. Even the words 'probiotic' and 'antioxidant' will disappear from food and food-supplement labels in the absence of confirmed specific health benefits.

To implement the regulations, the EFSA — whose role is to supply scientific advice in support of EU policies — set about drawing up a register of permitted health claims for food ingredients, asking its expert committees to recommend for inclusion only those for which the claims have been unambiguously proven in healthy populations.

Of the 2,927 consolidated health claims for different ingredients examined by the EFSA, only 241 passed muster. A second chance was given to 91 of the non-approved claims, 74 of which related to microorganisms — pre- and probiotics. But the EFSA rejected all of the resubmitted dossiers except two: prunes for normal bowel function and the polysaccharide α-cyclodextrin, a soluble dietary fibre, for limiting the rise in blood glucose after a meal.

The high rejection rate dismayed sections of the multibillion-euro health-food industry, which has lobbied fiercely against this legislation with disingenuous arguments that it inappropriately applies pharmaceutical standards to foods. In fact, it is the industry that has tried to make pharmaceutical-level claims for its products while bypassing medical-registration procedures and costly quality control.

But there is an elephant in the room: botanicals. Around 2,000 health claims for plant-based foods are on hold at the ESFA as manufacturers plead for special treatment. The manufacturers refer to the 2001 EU medicines directive which allowed traditional herbal medicines a simplified registration procedure not requiring rigorous proof of efficacy. The EU health commissioner must decide whether plant-based products marketed as health-promoting foods should be treated with similar leniency, but the decision has been shamefully delayed. The new commissioner, Tonio Borg, has been in office for only a few weeks, but he needs to bring clarity quickly — and to firmly reject moves to weaken requirements for scientific evidence.

"Already, many health-food companies, wary of possible litigation, have toned down their marketing claims."

The European Court of Justice may yet force a decision on the commission if it continues to prevaricate. The court is currently hearing three cases of unfair competition from manufacturers stopped from making claims on non-botanical food and food supplements. Yogurt manufacturers, for example, do not consider it fair that botanicals can continue to be marketed with unproven claims of improving immune defences when they are no longer allowed to do so.

Faced with the new legislation, the yogurt industry has in fact buckled down to generate the scientific evidence that the commission wants. Manufacturers have joined forces to produce a meta-analysis of published evidence on some probiotic strains to try to get at least some claims onto the permitted list before it circulates to parliament. And they are also launching double-blinded clinical studies to prove that certain microbial strains have particular effects on health. That is an expensive exercise, but it is the price that must be paid by those who want to stay in the game. ■

Realities of risk

We should focus on dangers that we can control, and particularly on those of our own creation.

When policy-makers consider global risks, they tend to extrapolate from headlines. Troubles in the eurozone could spiral out of control, conflict in Syria could spark wider unrest, the H5N1 bird-flu virus could mutate and spread from person to person in a global pandemic. Those dangers and many others are clear and present; society is well-advised to prepare for them, and takes good advice on how to do so. The World Economic Forum (WEF), for one, publishes an annual risk-assessment report.

There is, however, another category of risk: the unheralded dangers that sneak up on us. Many are the unforeseen consequences of progress, of humanity's scientific and technological quests. For its 2013 report, the WEF asked the editors and journalists of *Nature* to identify five of the most disruptive of these risks, dubbed 'X factors'.

Neuroscientists, for instance, are avidly pursuing drugs and devices that could deliver real cognitive enhancement — not just sharpening our alertness and ability to focus, as certain drugs already do, but upping our intellectual firepower. From students to business executives, the demand for such drugs would be huge — and so would the potential for an X factor to strike. Few drugs affect just a single target. The neurotransmitter systems important to cognition also serve other functions, raising the spectre of serious side effects: for example, a drug that boosts memory might also make the user more prone to impulsive behaviour. And then there are the ethical conundrums: should the market decide who gets the benefits of these drugs and who does not? Should they be banned, to level the playing field — or subsidized, for the same reason?

Climatologists have more mixed feelings about schemes for geoengineering — deliberately altering the climate system to combat the effects of rising greenhouse-gas emissions. In one scenario, high-flying jets or balloons would release a haze of sulphate particles into the stratosphere, dimming the Sun's rays and cooling the planet. Fearful that geoengineering could affect the climate system in unexpected ways, researchers have deliberated and studied it, and so far proposed only the most cautious of experiments. A second X factor looms here: geoengineering is simple and cheap enough that a rogue nation, or even a company, could deploy it on a large scale before its risks are well understood, perhaps triggering a widespread climatic crisis.

The other X factors that *Nature* staff identified are no less dramatic: the societal burden of the millions of people who, thanks to progress against killer diseases, will join the ranks of the disabled and those with dementia; catastrophic climate feedback such as the collapse of an ice sheet; and the possible social consequences of contact with alien life (although this is perhaps more X-files than X factor).

In this week's issue, *Nature* also takes up related topics. A World View on page 135 offers psychological insight into 'digital wildfires' (a danger explored in *Global Risks 2013*). These are a much more certain danger than the X factors — in fact, they are a regular feature of the Internet, flaring up practically by the minute as rumours or personal attacks race through cyberspace. Meanwhile, a Comment on page 157 examines the shortcomings of efforts to forecast state changes such as the collapse of ecosystems or the outbreak of epidemics.

Finally, a Feature on page 154 offers the liberating perspective that some of the very worst things that could happen — catastrophes that could devastate the biosphere, and human society with it — are out of our hands. These strokes of cosmic bad luck, among them supervolcanoes, apocalyptic tsunamis driven by undersea landslides, and mega solar flares, have all happened in the geological past and are sure to strike again someday. But in many cases there is little we can do to prepare or stave them off — and so there is no point in worrying. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

whenever new scientific evidence can be brought to bear — will be presented to the European Parliament for inspection next month, and adopted by member states later this year. Already, many health-food companies, wary of possible litigation, have toned down their marketing claims.

The regulations, drafted by the European Commission and adopted by the EU in 2007, are intended to put an end to the free ride that the food and food-supplement industries have enjoyed until now. Companies will no longer be able to market their products with unproven promises that consumers will be healthier, slimmer or have happier lives. Even the words 'probiotic' and 'antioxidant' will disappear from food and food-supplement labels in the absence of confirmed specific health benefits.

To implement the regulations, the EFSA — whose role is to supply scientific advice in support of EU policies — set about drawing up a register of permitted health claims for food ingredients, asking its expert committees to recommend for inclusion only those for which the claims have been unambiguously proven in healthy populations.

Of the 2,927 consolidated health claims for different ingredients examined by the EFSA, only 241 passed muster. A second chance was given to 91 of the non-approved claims, 74 of which related to microorganisms — pre- and probiotics. But the EFSA rejected all of the resubmitted dossiers except two: prunes for normal bowel function and the polysaccharide α-cyclodextrin, a soluble dietary fibre, for limiting the rise in blood glucose after a meal.

The high rejection rate dismayed sections of the multibillion-euro health-food industry, which has lobbied fiercely against this legislation with disingenuous arguments that it inappropriately applies pharmaceutical standards to foods. In fact, it is the industry that has tried to make pharmaceutical-level claims for its products while bypassing medical-registration procedures and costly quality control.

But there is an elephant in the room: botanicals. Around 2,000 health claims for plant-based foods are on hold at the ESFA as manufacturers plead for special treatment. The manufacturers refer to the 2001 EU medicines directive which allowed traditional herbal medicines a simplified registration procedure not requiring rigorous proof of efficacy. The EU health commissioner must decide whether plant-based products marketed as health-promoting foods should be treated with similar leniency, but the decision has been shamefully delayed. The new commissioner, Tonio Borg, has been in office for only a few weeks, but he needs to bring clarity quickly — and to firmly reject moves to weaken requirements for scientific evidence.

"Already, many health-food companies, wary of possible litigation, have toned down their marketing claims."

The European Court of Justice may yet force a decision on the commission if it continues to prevaricate. The court is currently hearing three cases of unfair competition from manufacturers stopped from making claims on non-botanical food and food supplements. Yogurt manufacturers, for example, do not consider it fair that botanicals can continue to be marketed with unproven claims of improving immune defences when they are no longer allowed to do so.

Faced with the new legislation, the yogurt industry has in fact buckled down to generate the scientific evidence that the commission wants. Manufacturers have joined forces to produce a meta-analysis of published evidence on some probiotic strains to try to get at least some claims onto the permitted list before it circulates to parliament. And they are also launching double-blinded clinical studies to prove that certain microbial strains have particular effects on health. That is an expensive exercise, but it is the price that must be paid by those who want to stay in the game. ■

Realities of risk

We should focus on dangers that we can control, and particularly on those of our own creation.

When policy-makers consider global risks, they tend to extrapolate from headlines. Troubles in the eurozone could spiral out of control, conflict in Syria could spark wider unrest, the H5N1 bird-flu virus could mutate and spread from person to person in a global pandemic. Those dangers and many others are clear and present; society is well-advised to prepare for them, and takes good advice on how to do so. The World Economic Forum (WEF), for one, publishes an annual risk-assessment report.

There is, however, another category of risk: the unheralded dangers that sneak up on us. Many are the unforeseen consequences of progress, of humanity's scientific and technological quests. For its 2013 report, the WEF asked the editors and journalists of *Nature* to identify five of the most disruptive of these risks, dubbed 'X factors'.

Neuroscientists, for instance, are avidly pursuing drugs and devices that could deliver real cognitive enhancement — not just sharpening our alertness and ability to focus, as certain drugs already do, but upping our intellectual firepower. From students to business executives, the demand for such drugs would be huge — and so would the potential for an X factor to strike. Few drugs affect just a single target. The neurotransmitter systems important to cognition also serve other functions, raising the spectre of serious side effects: for example, a drug that boosts memory might also make the user more prone to impulsive behaviour. And then there are the ethical conundrums: should the market decide who gets the benefits of these drugs and who does not? Should they be banned, to level the playing field — or subsidized, for the same reason?

Climatologists have more mixed feelings about schemes for geoengineering — deliberately altering the climate system to combat the effects of rising greenhouse-gas emissions. In one scenario, high-flying jets or balloons would release a haze of sulphate particles into the stratosphere, dimming the Sun's rays and cooling the planet. Fearful that geoengineering could affect the climate system in unexpected ways, researchers have deliberated and studied it, and so far proposed only the most cautious of experiments. A second X factor looms here: geoengineering is simple and cheap enough that a rogue nation, or even a company, could deploy it on a large scale before its risks are well understood, perhaps triggering a widespread climatic crisis.

The other X factors that *Nature* staff identified are no less dramatic: the societal burden of the millions of people who, thanks to progress against killer diseases, will join the ranks of the disabled and those with dementia; catastrophic climate feedback such as the collapse of an ice sheet; and the possible social consequences of contact with alien life (although this is perhaps more X-files than X factor).

In this week's issue, *Nature* also takes up related topics. A World View on page 135 offers psychological insight into 'digital wildfires' (a danger explored in *Global Risks 2013*). These are a much more certain danger than the X factors — in fact, they are a regular feature of the Internet, flaring up practically by the minute as rumours or personal attacks race through cyberspace. Meanwhile, a Comment on page 157 examines the shortcomings of efforts to forecast state changes such as the collapse of ecosystems or the outbreak of epidemics.

Finally, a Feature on page 154 offers the liberating perspective that some of the very worst things that could happen — catastrophes that could devastate the biosphere, and human society with it — are out of our hands. These strokes of cosmic bad luck, among them super-volcanoes, apocalyptic tsunamis driven by undersea landslides, and mega solar flares, have all happened in the geological past and are sure to strike again someday. But in many cases there is little we can do to prepare or stave them off — and so there is no point in worrying. ■

NATURE.COM
To comment online, click on Editorials at: go.nature.com/xhunqv

WORLD VIEW *A personal take on events*

ETC/RIT

Rumour research can douse digital wildfires

Work on how rumours arise and spread could help to dampen the effects of damaging misinformation circulating on the Internet, says **Nicholas DiFonzo**.

The metaphor 'digital wildfire' — one of numerous threats to society identified this week by the World Economic Forum in its *Global Risks* report — is well chosen. The image highlights the risks of the rapid spread of uncontrollable and destructive information — risks that are made possible by an ever-growing worldwide web of communication channels.

Many digital wildfires start as rumours, and social psychologists have long studied how rumours spread. The results of this research can therefore help in the prevention of and response to digital wildfires.

Why is this important? A digital wildfire ignited by a humorous complaint about United Airlines posted on YouTube by musician Dave Carroll saw the company's stock drop by 10%. A false Internet rumour that Syrian President Bashar al-Assad had been killed caused oil prices to rise. These are not trivial matters.

Research on rumour has come up with four concepts that may be applicable to digital wildfires: motivation, situation, narrative context and trust.

The spreading of rumours is an act of goal-oriented communication, often motivated by a desire to find out whether they are true. Other common motives relate to the sender–receiver relationship: to please or entertain the receiver and to make the receiver like or respect the sender. Self-oriented aims include: fulfilment of the sender's wishes; the boosting of self-esteem (or defence against threats to one's sense of self, cherished values or group identity); the exacting of revenge; and the achievement of a strategic goal (as with propaganda).

Situation matters, too. Just as forests are especially vulnerable to fire during drought, so too some situations make digital networks susceptible to wildfires. Ambiguity and threats to assets, both tangible (life, health or wealth) and psychological (group honour or cherished values) drive fact-finding rumours. Situations of conflict or rivalry produce strong defensive sentiments that welcome negative rumours about a rival group and positive rumours about our own. Perceived injustice often prompts people to take revenge.

Rumours reflect and resonate with larger cultural, historical and ideological narrative contexts, which render the claims more plausible and understandable to some groups than others. A YouTube trailer that defamed the Prophet Muhammad as a murderer, womanizer, bisexual and paedophile resonated with the widespread Islamic narrative of Westerners as desecrating, condescending and shameless barbarians who derogate the sacred and threaten moral order; violent reactions to it throughout the Muslim world cannot be understood without taking account of this context.

The most powerful ingredient, however, could be trust. Greater trust dampens perceptions of threat, makes negative rumours less believable and enhances the effectiveness of rumour refutation or response. Research has shown that, under conditions of interdependence and equal status, increased contact with members of rival groups decreases negative stereotyping and increases trust.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/bhdnnb

But in the digital world, rival groups tend to have little contact with each other. Recruiting members from opposing factions to projects with shared goals may therefore help to ease these tensions. In my own laboratory, when groups of strong Republicans and Democrats communicated online to accomplish a task, fewer believed polarizing rumours such as "Democrats abuse drugs" and "Republicans are racist".

These strands of rumour psychology can help us to understand digital wildfires. Why did Carroll's YouTube complaint that United Airlines broke his US$3,500 guitar, then refused to pay for repairs go viral? First, Carroll is a professional singer and his video was clever, humorous and entertaining; passing it along to friends and co-workers enhanced sender–receiver friendships.

GREATER TRUST DAMPENS PERCEPTIONS OF THREAT.

Second, he tapped into a wellspring of distrust: within two weeks he received 10,000 e-mails from people asking for help with other customer-service problems. Anyone who has ever seen baggage lost or damaged can understand how this wildfire was fuelled by a widespread sense of powerlessness, frustration and dissatisfaction with an ever-more-Spartan post-9/11 airline industry. It drew on an implicit cultural narrative of David versus Corporate Goliath, and offered a hard-to-resist opportunity to exact revenge.

But United's response — which included reimbursement offers and a manager's apology — failed to stem the epidemic. More potent medicine was needed. The company could have, for example, come up with a country-music style video in which its president offered a no-excuse apology and announced measures to revamp the complaint department and to ensure safe baggage handling. Such a strategy would have removed revenge motivation (through gestures of justice and respect) and defused the situation (frustrations about baggage safety). It would also have conveyed benign intent (trust) and invoked the common narrative of apology–restitution–forgiveness. It might even have been entertaining to pass along.

Virgin Atlantic responded much more appropriately to a complaint about in-flight food that was starting to spread widely. The airline's owner immediately phoned the passenger to apologize and offered him the opportunity to assist in the airline's food selection. Virgin dampened the revenge motive by offering a high-quality public apology and restitution, and gave power to the offended party; because of the apparent humility and sincerity of these actions, Virgin did much to engender trust. The wildfire was doused. ■ SEE EDITORIAL P.134 AND FEATURE P.154

Nicholas DiFonzo *is professor of psychology at the Rochester Institute of Technology in New York and author of* The Watercooler Effect.
e-mail: nick@profnick.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

CLIMATE SCIENCE

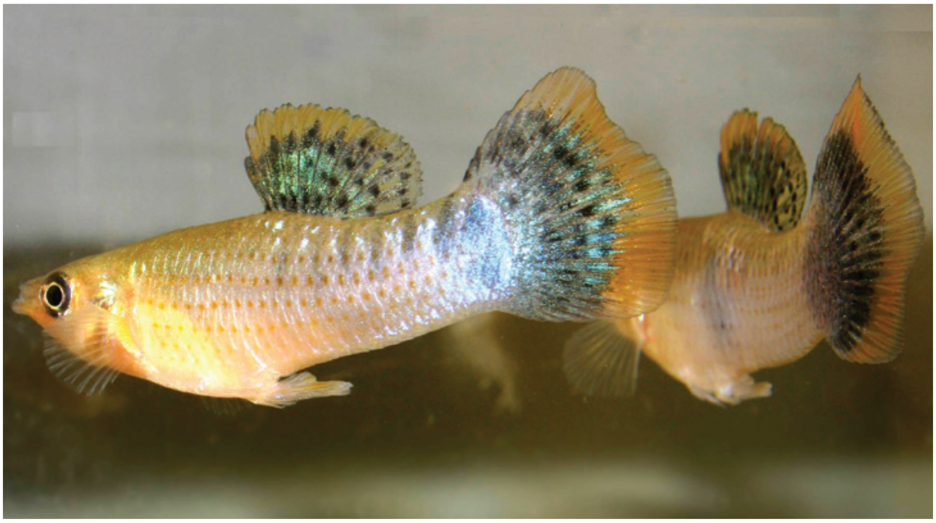
West Antarctic warming hotspot

West Antarctica warmed by about 2.4 °C between 1958 and 2010, making it one of the fastest-warming areas of the planet.

Previously reported temperature trends in West Antarctica have been disputed, and records in the region are sparse. David Bromwich at Ohio State University in Columbus and his colleagues analysed the most complete set of records available, from the Byrd Station US research outpost, and filled in missing observations with modelling and data from other sites. Their analysis shows warming in West Antarctica in the spring, summer and winter.

Despite the lack of evidence for increased surface melting in this region so far, a continued rise in summer temperatures could boost the chances of more frequent and intense melting events, the authors say.

Nature Geosci. **http://dx.doi.org/10.1038/ngeo1671 (2012)**



D. BIERBACH

ANIMAL BEHAVIOUR

Bisexuality boosts attractiveness

Female fish find some males more attractive if they have seen the males engaging in mating behaviour, even when such behaviour was with other males.

David Bierbach and his colleagues at the University of Frankfurt in Germany assessed the preferences of female Atlantic mollies (*Poecilia mexicana*) for various males of the same species (pictured) by showing them video animations of the males and measuring the time they spent in proximity to the images. Females preferred colourful males at first, but showed an increased preference for drabber males that they witnessed engaging in either homosexual or heterosexual mating behaviour. Another experiment confirmed that the females could tell the difference between male and female fish in the animations.

The authors suggest that this increased female preference for sexually active males may explain the prevalence of bisexuality in some group-living species. Homosexual behaviour, they say, confers a reproductive advantage on males of those species by increasing their likelihood of future heterosexual mating.

Biol. Lett. 9, **20121038 (2013)**

GENOMICS

Single-cell sequencing

Sequencing a genome usually requires DNA from thousands or even millions of cells, but a technique now allows more than 90% of the genome of a single cell to be sequenced. This could enable studies, for example, on how mutated cancer cells emerge or how individual neurons differ.

Sunney Xie at Harvard University in Cambridge, Massachusetts, and his colleagues developed the method, which uses short DNA molecules called primers. These molecules are added to DNA that has been isolated from a single cell, and that stick to its strands to act as starting points for DNA replication. The primers are designed to reduce the excessive copying of some portions of the genome at the expense of others — a problem that has plagued other attempts at single-cell sequencing.

Science 338, **1622–1626 (2012)**

For a longer story on this research, see http://go.nature.com/fdvwb2

ASTRONOMY

Most powerful quasar yet

Astronomers have discovered a quasar that is expelling material 10 times more powerfully than any previously observed.

The quasar — the luminous centre of a galaxy powered by a supermassive black hole — each year blasts out material equivalent to the mass of 400 Suns. Benoit Borguet at Virginia Polytechnic Institute and State University in Blacksburg and his colleagues spotted the object, dubbed SDSS J1106+1939, in April 2011 using the Very Large Telescope, a European Southern Observatory facility on Cerro Paranal in Chile. The quasar's existence helps to strengthen models of galactic evolution, which use outflows of supermassive black holes as a feedback mechanism to constrain star formation.

Astrophys J. 762, **49 (2013)**

RESEARCH HIGHLIGHTS

Selections from the scientific literature

CLIMATE SCIENCE

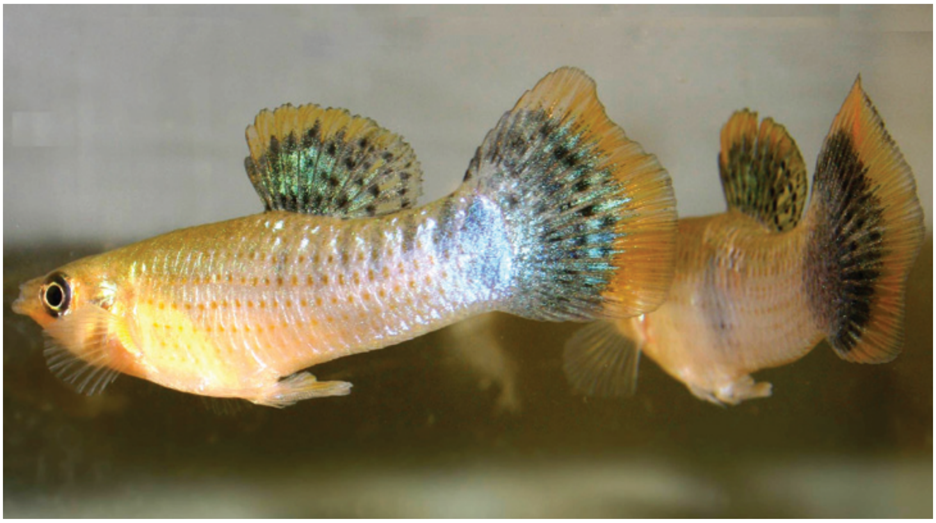
West Antarctic warming hotspot

West Antarctica warmed by about 2.4 °C between 1958 and 2010, making it one of the fastest-warming areas of the planet.

Previously reported temperature trends in West Antarctica have been disputed, and records in the region are sparse. David Bromwich at Ohio State University in Columbus and his colleagues analysed the most complete set of records available, from the Byrd Station US research outpost, and filled in missing observations with modelling and data from other sites. Their analysis shows warming in West Antarctica in the spring, summer and winter.

Despite the lack of evidence for increased surface melting in this region so far, a continued rise in summer temperatures could boost the chances of more frequent and intense melting events, the authors say.

Nature Geosci. **http://dx.doi.org/10.1038/ngeo1671 (2012)**



D. BIERBACH

ANIMAL BEHAVIOUR

Bisexuality boosts attractiveness

Female fish find some males more attractive if they have seen the males engaging in mating behaviour, even when such behaviour was with other males.

David Bierbach and his colleagues at the University of Frankfurt in Germany assessed the preferences of female Atlantic mollies (*Poecilia mexicana*) for various males of the same species (pictured) by showing them video animations of the males and measuring the time they spent in proximity to the images. Females preferred colourful males at first, but showed an increased preference for drabber males that they witnessed engaging in either homosexual or heterosexual mating behaviour. Another experiment confirmed that the females could tell the difference between male and female fish in the animations.

The authors suggest that this increased female preference for sexually active males may explain the prevalence of bisexuality in some group-living species. Homosexual behaviour, they say, confers a reproductive advantage on males of those species by increasing their likelihood of future heterosexual mating.

Biol. Lett. 9, **20121038 (2013)**

GENOMICS

Single-cell sequencing

Sequencing a genome usually requires DNA from thousands or even millions of cells, but a technique now allows more than 90% of the genome of a single cell to be sequenced. This could enable studies, for example, on how mutated cancer cells emerge or how individual neurons differ.

Sunney Xie at Harvard University in Cambridge, Massachusetts, and his colleagues developed the method, which uses short DNA molecules called primers. These molecules are added to DNA that has been isolated from a single cell, and that stick to its strands to act as starting points for DNA replication. The primers are designed to reduce the excessive copying of some portions of the genome at the expense of others — a problem that has plagued other attempts at single-cell sequencing.

Science 338, **1622–1626 (2012)**

For a longer story on this research, see http://go.nature.com/fdvwb2

ASTRONOMY

Most powerful quasar yet

Astronomers have discovered a quasar that is expelling material 10 times more powerfully than any previously observed.

The quasar — the luminous centre of a galaxy powered by a supermassive black hole — each year blasts out material equivalent to the mass of 400 Suns. Benoit Borguet at Virginia Polytechnic Institute and State University in Blacksburg and his colleagues spotted the object, dubbed SDSS J1106+1939, in April 2011 using the Very Large Telescope, a European Southern Observatory facility on Cerro Paranal in Chile. The quasar's existence helps to strengthen models of galactic evolution, which use outflows of supermassive black holes as a feedback mechanism to constrain star formation.

Astrophys J. 762, **49 (2013)**

ASTRONOMY

A galaxy far, far away

Using the Hubble Space Telescope, astronomers have detected light from the most distant object yet found — a fledgling galaxy that existed when the Universe was just over 420 million years old. The discovery shows that stars were organizing into galaxies when the Universe was just 3% of its current age of 13.7 billion years. Dan Coe of the Space Telescope Science Institute in Baltimore, Maryland, and his colleagues spotted light from the galaxy by taking advantage of an intervening lens of dark matter that magnified the object. Coe says that this and other discoveries suggest that light from galaxies might be responsible for ionizing the early Universe.

Astrophys J. 762, **32 (2013)**

NEUROPROSTHETICS

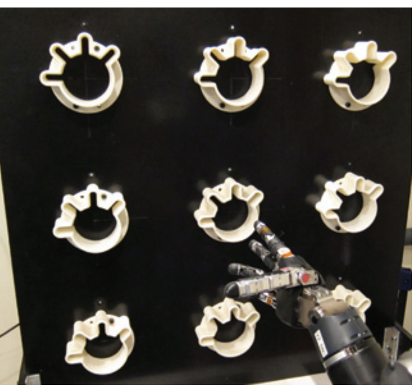
Fine mind control of machine

A woman paralysed in all four limbs has shown that she can move a prosthetic arm with unprecedented precision using only her brain to control the arm.

Andrew Schwartz at the University of Pittsburgh in Pennsylvania and his colleagues implanted two microelectrodes into the woman's motor cortex, the part of the brain responsible for initiating movement. The devices recorded cortical activity associated with her desire to move the robotic arm. An external decoder connected to the electrodes and the arm converted these signals into electronic instructions for the limb.

The woman began training to use the arm (**pictured**) — which has wrist-joint movement comparable to that of a human hand — less than two weeks after implantation. Within two days she could make simple movements, and within 13 weeks she could grasp and move objects almost as fast, smoothly and accurately as an able-bodied person.

Lancet **http://dx.doi.org/10.1016/S0140-6736(12)61816-9 (2012)**



THE LANCET/ELSEVIER

GENETIC ENGINEERING

Precise genomic editing

A set of genetic tools based on components of a bacterial immune system can make the process of adding, deleting or replacing specific genes in model organisms and human cells easier and more precise.

Current gene-editing enzymes are expensive and difficult to use. Feng Zhang at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, and his colleagues developed an alternative method focusing on an enzyme, called Cas9, and two RNA molecules from a bacterial immune system known as CRISPR. These molecules work together to recognize and cut viral DNA. The team modified these molecules so that they could be used to edit human and mouse genomes at precise and even multiple locations.

George Church at Harvard Medical School in Boston, Massachusetts, and his colleagues showed that their CRISPR-based gene-editing system could be used in human cells, including reprogrammed stem cells.

Science **http://dx.doi.org/10.1126/science.1231143; http://dx.doi.org/10.1126/science.1232033 (2013)**

COMMUNITY CHOICE

The most viewed papers in science

MOLECULAR IMAGING

Direct images of DNA

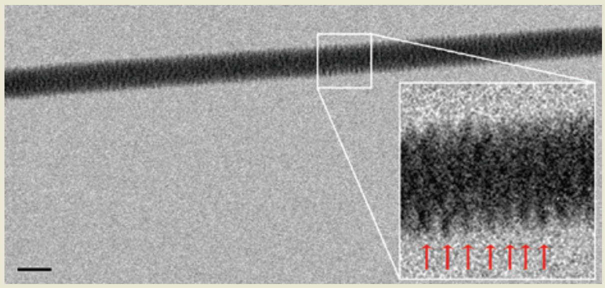
HIGHLY READ on pubs.acs.org in December

Nearly 60 years after the structure of DNA was first discovered, scientists using transmission electron microscopy have captured direct images of it.

Enzo di Fabrizio at the Italian Institute of Technology in Genoa and his colleagues spread droplets of fluid containing DNA from a bacterium-infecting virus onto silicon wafers that were etched with micrometre-scale pillars and holes. As the droplets dried overnight, the DNA stretched taut across the tops of the pillars, passing over some of the holes. The researchers then imaged the DNA (**pictured**; scale bar, 20 nanometres) by beaming electrons through the holes.

The team observed a structure with helices that repeated roughly every 2.7 nanometres (inset, red arrows), consistent with A-DNA — a non-physiological DNA conformation known to occur in dehydrated samples.

Nano Lett. 12, **6453–6458 (2012)**



AM. CHEM. SOC.

STEM CELLS

Immune cells rejuvenated

Two groups in Japan have reprogrammed cells that fight HIV and cancer, boosting the number and lifespan of the cells.

The body's immune system makes T cells that target viruses and cancer, but the pool of these cells quickly becomes exhausted. Shin Kaneko and Hiromitsu Nakauchi at the University of Tokyo and their colleagues took a group of HIV-specific T cells from a person with HIV. They reprogrammed the cells into stem cells, and then added key growth factors so that the stem cells developed into the original type of T cell. The rejuvenated cells retained their ability to target HIV, had an enhanced capacity for proliferation and had longer telomeres, or caps on the end of their chromosomes, suggesting a longer lifespan.

Hiroshi Kawamoto at the RIKEN Research Center for Allergy and Immunology in Yokohama and his colleagues used a similar approach to reprogram T cells specific to melanoma. The reprogrammed T cells could still recognize a protein commonly found on the surface of melanoma cells, and were present in greater numbers than they had been in the original population.

Cell Stem Cell 12, **114–126; 31–36 (2013)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

ASTRONOMY

A galaxy far, far away

Using the Hubble Space Telescope, astronomers have detected light from the most distant object yet found — a fledgling galaxy that existed when the Universe was just over 420 million years old. The discovery shows that stars were organizing into galaxies when the Universe was just 3% of its current age of 13.7 billion years. Dan Coe of the Space Telescope Science Institute in Baltimore, Maryland, and his colleagues spotted light from the galaxy by taking advantage of an intervening lens of dark matter that magnified the object. Coe says that this and other discoveries suggest that light from galaxies might be responsible for ionizing the early Universe.

Astrophys J. **762, 32 (2013)**

NEUROPROSTHETICS

Fine mind control of machine

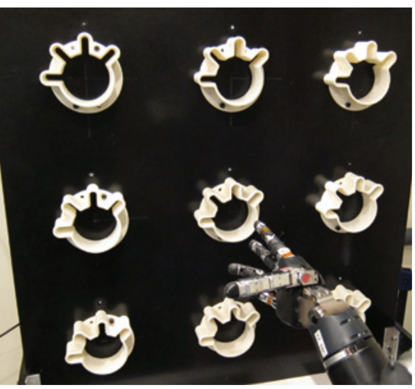
A woman paralysed in all four limbs has shown that she can move a prosthetic arm with unprecedented precision using only her brain to control the arm.

Andrew Schwartz at the University of Pittsburgh in Pennsylvania and his colleagues implanted two microelectrodes into the woman's motor cortex, the part of the brain responsible for initiating movement. The devices recorded cortical activity associated with her desire to move the robotic arm. An external decoder connected to the electrodes and the arm converted these signals into electronic instructions for the limb.

The woman began training to use the arm (**pictured**) — which has wrist-joint movement comparable to that of a human hand — less than two weeks after implantation. Within two days she could make simple movements, and within 13 weeks she could grasp and move objects almost as fast, smoothly and accurately as an able-bodied person.

Lancet **http://dx.doi.org/10.1016/S0140-6736(12)61816-9 (2012)**

THE LANCET/ELSEVIER



GENETIC ENGINEERING

Precise genomic editing

A set of genetic tools based on components of a bacterial immune system can make the process of adding, deleting or replacing specific genes in model organisms and human cells easier and more precise.

Current gene-editing enzymes are expensive and difficult to use. Feng Zhang at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, and his colleagues developed an alternative method focusing on an enzyme, called Cas9, and two RNA molecules from a bacterial immune system known as CRISPR. These molecules work together to recognize and cut viral DNA. The team modified these molecules so that they could be used to edit human and mouse genomes at precise and even multiple locations.

George Church at Harvard Medical School in Boston, Massachusetts, and his colleagues showed that their CRISPR-based gene-editing system could be used in human cells, including reprogrammed stem cells.

Science **http://dx.doi.org/10.1126/science.1231143; http://dx.doi.org/10.1126/science.1232033 (2013)**

COMMUNITY CHOICE

The most viewed papers in science

MOLECULAR IMAGING

Direct images of DNA

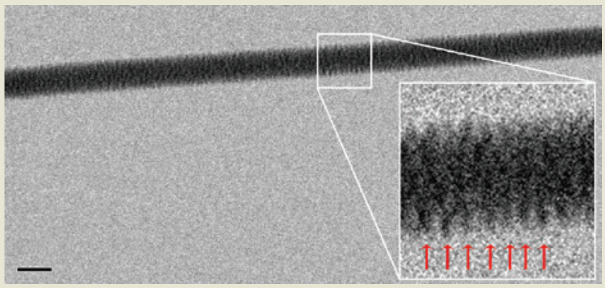
HIGHLY READ on pubs.acs.org in December

Nearly 60 years after the structure of DNA was first discovered, scientists using transmission electron microscopy have captured direct images of it.

Enzo di Fabrizio at the Italian Institute of Technology in Genoa and his colleagues spread droplets of fluid containing DNA from a bacterium-infecting virus onto silicon wafers that were etched with micrometre-scale pillars and holes. As the droplets dried overnight, the DNA stretched taut across the tops of the pillars, passing over some of the holes. The researchers then imaged the DNA (**pictured**; scale bar, 20 nanometres) by beaming electrons through the holes.

The team observed a structure with helices that repeated roughly every 2.7 nanometres (inset, red arrows), consistent with A-DNA — a non-physiological DNA conformation known to occur in dehydrated samples.

Nano Lett. 12, **6453–6458 (2012)**



AM. CHEM. SOC.

STEM CELLS

Immune cells rejuvenated

Two groups in Japan have reprogrammed cells that fight HIV and cancer, boosting the number and lifespan of the cells.

The body's immune system makes T cells that target viruses and cancer, but the pool of these cells quickly becomes exhausted. Shin Kaneko and Hiromitsu Nakauchi at the University of Tokyo and their colleagues took a group of HIV-specific T cells from a person with HIV. They reprogrammed the cells into stem cells, and then added key growth factors so that the stem cells developed into the original type of T cell. The rejuvenated cells retained their ability to target HIV, had an enhanced capacity for proliferation and had longer telomeres, or caps on the end of their chromosomes, suggesting a longer lifespan.

Hiroshi Kawamoto at the RIKEN Research Center for Allergy and Immunology in Yokohama and his colleagues used a similar approach to reprogram T cells specific to melanoma. The reprogrammed T cells could still recognize a protein commonly found on the surface of melanoma cells, and were present in greater numbers than they had been in the original population.

Cell Stem Cell 12, **114–126; 31–36 (2013)**

ASTRONOMY

A galaxy far, far away

Using the Hubble Space Telescope, astronomers have detected light from the most distant object yet found — a fledgling galaxy that existed when the Universe was just over 420 million years old. The discovery shows that stars were organizing into galaxies when the Universe was just 3% of its current age of 13.7 billion years. Dan Coe of the Space Telescope Science Institute in Baltimore, Maryland, and his colleagues spotted light from the galaxy by taking advantage of an intervening lens of dark matter that magnified the object. Coe says that this and other discoveries suggest that light from galaxies might be responsible for ionizing the early Universe.

Astrophys J. 762, **32 (2013)**

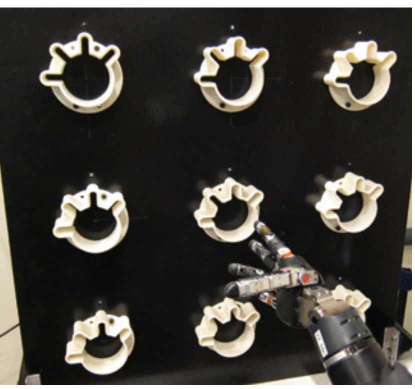
NEUROPROSTHETICS

Fine mind control of machine

A woman paralysed in all four limbs has shown that she can move a prosthetic arm with unprecedented precision using only her brain to control the arm.

Andrew Schwartz at the University of Pittsburgh in Pennsylvania and his colleagues implanted two microelectrodes into the woman's motor cortex, the part of the brain responsible for initiating movement. The devices recorded cortical activity associated with her desire to move the robotic arm. An external decoder connected to the electrodes and the arm converted these signals into electronic instructions for the limb.

THE LANCET/ELSEVIER



The woman began training to use the arm (**pictured**) — which has wrist-joint movement comparable to that of a human hand — less than two weeks after implantation. Within two days she could make simple movements, and within 13 weeks she could grasp and move objects almost as fast, smoothly and accurately as an able-bodied person.

Lancet **http://dx.doi.org/10.1016/S0140-6736(12)61816-9 (2012)**

GENETIC ENGINEERING

Precise genomic editing

A set of genetic tools based on components of a bacterial immune system can make the process of adding, deleting or replacing specific genes in model organisms and human cells easier and more precise.

Current gene-editing enzymes are expensive and difficult to use. Feng Zhang at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, and his colleagues developed an alternative method focusing on an enzyme, called Cas9, and two RNA molecules from a bacterial immune system known as CRISPR. These molecules work together to recognize and cut viral DNA. The team modified these molecules so that they could be used to edit human and mouse genomes at precise and even multiple locations.

George Church at Harvard Medical School in Boston, Massachusetts, and his colleagues showed that their CRISPR-based gene-editing system could be used in human cells, including reprogrammed stem cells.

Science **http://dx.doi.org/10.1126/science.1231143; http://dx.doi.org/10.1126/science.1232033 (2013)**

COMMUNITY CHOICE

The most viewed papers in science

MOLECULAR IMAGING

Direct images of DNA

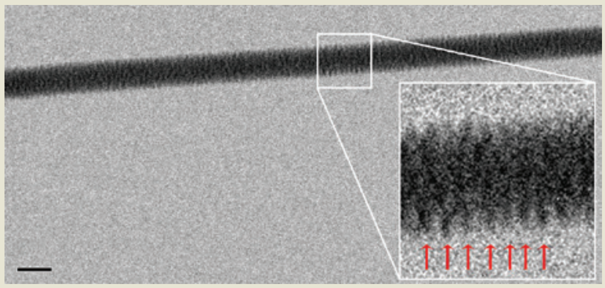
HIGHLY READ on pubs.acs.org in December

Nearly 60 years after the structure of DNA was first discovered, scientists using transmission electron microscopy have captured direct images of it.

Enzo di Fabrizio at the Italian Institute of Technology in Genoa and his colleagues spread droplets of fluid containing DNA from a bacterium-infecting virus onto silicon wafers that were etched with micrometre-scale pillars and holes. As the droplets dried overnight, the DNA stretched taut across the tops of the pillars, passing over some of the holes. The researchers then imaged the DNA (**pictured**; scale bar, 20 nanometres) by beaming electrons through the holes.

The team observed a structure with helices that repeated roughly every 2.7 nanometres (inset, red arrows), consistent with A-DNA — a non-physiological DNA conformation known to occur in dehydrated samples.

Nano Lett. 12, **6453–6458 (2012)**



AM. CHEM. SOC.

STEM CELLS

Immune cells rejuvenated

Two groups in Japan have reprogrammed cells that fight HIV and cancer, boosting the number and lifespan of the cells.

The body's immune system makes T cells that target viruses and cancer, but the pool of these cells quickly becomes exhausted. Shin Kaneko and Hiromitsu Nakauchi at the University of Tokyo and their colleagues took a group of HIV-specific T cells from a person with HIV. They reprogrammed the cells into stem cells, and then added key growth factors so that the stem cells developed into the original type of T cell. The rejuvenated cells retained their ability to target HIV, had an enhanced capacity for proliferation and had longer telomeres, or caps on the end of their chromosomes, suggesting a longer lifespan.

Hiroshi Kawamoto at the RIKEN Research Center for Allergy and Immunology in Yokohama and his colleagues used a similar approach to reprogram T cells specific to melanoma. The reprogrammed T cells could still recognize a protein commonly found on the surface of melanoma cells, and were present in greater numbers than they had been in the original population.

Cell Stem Cell 12, **114–126; 31–36 (2013)**

NATURE.COM

For the latest research published by *Nature* visit:

www.nature.com/latestresearch

ASTRONOMY

A galaxy far, far away

Using the Hubble Space Telescope, astronomers have detected light from the most distant object yet found — a fledgling galaxy that existed when the Universe was just over 420 million years old. The discovery shows that stars were organizing into galaxies when the Universe was just 3% of its current age of 13.7 billion years. Dan Coe of the Space Telescope Science Institute in Baltimore, Maryland, and his colleagues spotted light from the galaxy by taking advantage of an intervening lens of dark matter that magnified the object. Coe says that this and other discoveries suggest that light from galaxies might be responsible for ionizing the early Universe.

Astrophys J. 762, **32 (2013)**

NEUROPROSTHETICS

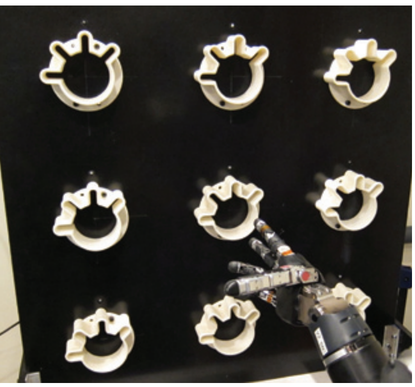
Fine mind control of machine

A woman paralysed in all four limbs has shown that she can move a prosthetic arm with unprecedented precision using only her brain to control the arm.

Andrew Schwartz at the University of Pittsburgh in Pennsylvania and his colleagues implanted two microelectrodes into the woman's motor cortex, the part of the brain responsible for initiating movement. The devices recorded cortical activity associated with her desire to move the robotic arm. An external decoder connected to the electrodes and the arm converted these signals into electronic instructions for the limb.

The woman began training to use the arm (**pictured**) — which has wrist-joint movement comparable to that of a human hand — less than two weeks after implantation. Within two days she could make simple movements, and within 13 weeks she could grasp and move objects almost as fast, smoothly and accurately as an able-bodied person.

Lancet **http://dx.doi.org/10.1016/S0140-6736(12)61816-9 (2012)**



THE LANCET/ELSEVIER

GENETIC ENGINEERING

Precise genomic editing

A set of genetic tools based on components of a bacterial immune system can make the process of adding, deleting or replacing specific genes in model organisms and human cells easier and more precise.

Current gene-editing enzymes are expensive and difficult to use. Feng Zhang at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, and his colleagues developed an alternative method focusing on an enzyme, called Cas9, and two RNA molecules from a bacterial immune system known as CRISPR. These molecules work together to recognize and cut viral DNA. The team modified these molecules so that they could be used to edit human and mouse genomes at precise and even multiple locations.

George Church at Harvard Medical School in Boston, Massachusetts, and his colleagues showed that their CRISPR-based gene-editing system could be used in human cells, including reprogrammed stem cells.

Science **http://dx.doi.org/10.1126/science.1231143; http://dx.doi.org/10.1126/science.1232033 (2013)**

COMMUNITY CHOICE

The most viewed papers in science

MOLECULAR IMAGING

Direct images of DNA

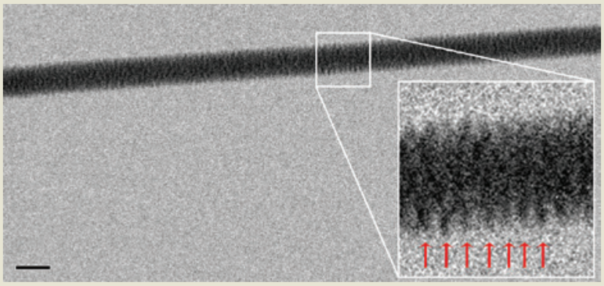
HIGHLY READ on pubs.acs.org in December

Nearly 60 years after the structure of DNA was first discovered, scientists using transmission electron microscopy have captured direct images of it.

Enzo di Fabrizio at the Italian Institute of Technology in Genoa and his colleagues spread droplets of fluid containing DNA from a bacterium-infecting virus onto silicon wafers that were etched with micrometre-scale pillars and holes. As the droplets dried overnight, the DNA stretched taut across the tops of the pillars, passing over some of the holes. The researchers then imaged the DNA (**pictured**; scale bar, 20 nanometres) by beaming electrons through the holes.

The team observed a structure with helices that repeated roughly every 2.7 nanometres (inset, red arrows), consistent with A-DNA — a non-physiological DNA conformation known to occur in dehydrated samples.

Nano Lett. 12, **6453–6458 (2012)**



AM. CHEM. SOC.

STEM CELLS

Immune cells rejuvenated

Two groups in Japan have reprogrammed cells that fight HIV and cancer, boosting the number and lifespan of the cells.

The body's immune system makes T cells that target viruses and cancer, but the pool of these cells quickly becomes exhausted. Shin Kaneko and Hiromitsu Nakauchi at the University of Tokyo and their colleagues took a group of HIV-specific T cells from a person with HIV. They reprogrammed the cells into stem cells, and then added key growth factors so that the stem cells developed into the original type of T cell. The rejuvenated cells retained their ability to target HIV, had an enhanced capacity for proliferation and had longer telomeres, or caps on the end of their chromosomes, suggesting a longer lifespan.

Hiroshi Kawamoto at the RIKEN Research Center for Allergy and Immunology in Yokohama and his colleagues used a similar approach to reprogram T cells specific to melanoma. The reprogrammed T cells could still recognize a protein commonly found on the surface of melanoma cells, and were present in greater numbers than they had been in the original population.

Cell Stem Cell 12, **114–126; 31–36 (2013)**

NATURE.COM

For the latest research published by *Nature* visit:

www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Stem-cell finale

The US Supreme Court has guaranteed that government-funded researchers will continue to be able to work with human embryonic stem cells. In declining to take up an appeal by the plaintiffs in the case of *Sherley* v. *Sebelius,* the high court ended a 3.5-year legal battle in which two researchers working with adult stem cells challenged the legality of the National Institutes of Health (NIH) funding the work, which requires the destruction of days-old embryos. NIH director Francis Collins said he was "very pleased" with the 7 January decision. See go.nature.com/ouwpo2 for more.

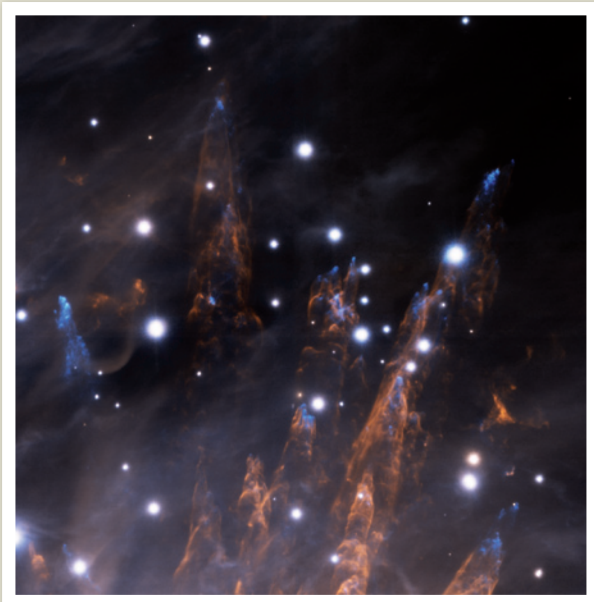
Poland GM ban

The Polish government on 2 January imposed a ban on the cultivation of two genetically modified (GM) crops: MON 810, an insect-resistant maize (corn); and Amflora, a starchy potato used in the paper industry. They are the only GM crops approved by European Union (EU) science advisory committees as safe for agriculture. Poland is one of nine EU members that have sought to impose such bans — despite EU legislation that requires all member states to permit cultivation of approved crops. European courts have ruled against bans in two countries. Many of the bans rely on a legal loophole, with countries claiming that there is new information on harm caused by the crops that is yet to be evaluated.

NUMBER CRUNCH

17 bn

Number of stars in the Milky Way that harbour a roughly Earth-sized planet in a close orbit, according to researchers at the Harvard–Smithsonian Center for Astrophysics. They presented their study on 7 January at the American Astronomical Society meeting in Long Beach, California.



GEMINI OBSERVATORY/AURA

Orion bullets shot in detail

The Gemini South telescope in Chile has been fitted with five lasers and three deformable mirrors, allowing astronomers to correct for atmospheric distortions over an exceptionally large field of view. On 9 January, at a meeting of the American Astronomical Society in Long Beach, California, astronomers unveiled the telescope's ultra-sharp portrait of the 'bullets' of gas seen in the Orion Nebula. In the image, clumps of iron gas (blue) race through the nebula, leaving behind pillars of hot, glowing hydrogen gas (orange).

Energy tax credits

The US alternative-energy industry received a bundle of tax incentives in the budget deal that was passed by Congress on 1 January to avert sweeping cuts in government spending. The deal offered an extra year for electricity producers to claim 10 years of tax credits for wind, geothermal and biomass projects. Lawmakers also reinstated a tax credit of US$1 per gallon (26.5 cents per litre) for biodiesel production, applied retroactively to 2012 when the credit expired, and extending through 2013.

Indian aspirations

India is setting its sights on becoming one of the top five global scientific powers by 2020, and should double its research expenditure to 2% of its gross domestic product in five years, according to prime minister Manmohan Singh. He was speaking as a new policy on science, technology and innovation was unveiled at the opening of the Indian Science Congress in Kolkata on 3 January. The policy is aimed at increasing the number of scientists in India by 66% by 2017, and at enhancing private-sector participation in research. Some of its aims, including spending targets, were proposed in India's Twelfth Five Year Plan, released on 27 December 2012. See go.nature.com/8te55t for more.

US food safety

On 4 January, the US Food and Drug Administration released the first of its long-awaited proposals for protecting food safety. The draft regulations are a result of the Food Safety Modernization Act, signed into law in January 2011, and are intended to shift the agency to a preventive, rather than reactive, stance on food-borne illness. One of the two proposed regulations requires food makers to develop plans for preventing contamination; the other sets safety standards for farms that grow produce, including, for example, permitted levels of microbes in irrigation water.

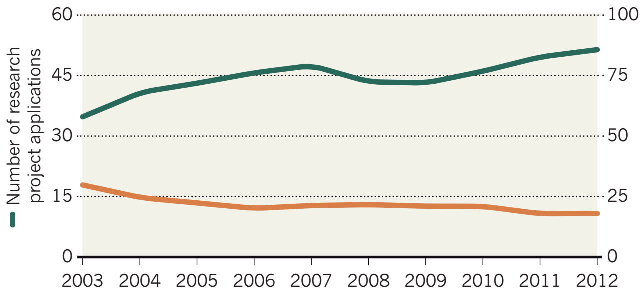
PEOPLE

Whistleblower out

A scientist who set up a whistle-blowing website to flag up problems in the scientific literature revealed his identity last week, after he shut down the site following legal threats. Paul Brookes, a biologist at the University of Rochester, New York, said he had registered www.science-fraud.org in July 2012, motivated by frustration at the channels available for dealing with scientific misconduct.

SOURCE: NIH

LONG ODDS AT THE NIH



RESEARCH

Success rates stuck

The success rate for grant applicants to the US National Institutes of Health (NIH) remained unchanged at a historic low of 18% in 2012 (see **graph**), the agency announced on 2 January. Sally Rockey, the NIH's deputy director for extramural research, wrote on her Rock Talk blog that the rate for research project grants was the same as that in the 2011 fiscal year, despite a 3.5% increase in applications to 51,313, a record high. Average grant size increased by 1.1%, to US$454,588. In all, the agency spent $15.9 billion of its $30.9-billion budget on research grants.

Boston pathogen lab

The US National Institutes of Health (NIH) will support a plan for a laboratory in Boston, Massachusetts, to work with some of the world's most dangerous pathogens. On 2 January, the NIH announced its final determination that Boston University's National Emerging Infectious Diseases Laboratory poses little risk to the surrounding community. The laboratory, which had been under review at NIH since 2001, must win approvals from state and local authorities before it can upgrade its research to biosafety level 4, the highest level of containment.

BUSINESS

Pharma split

Abbott Laboratories has spun off its drug-discovery business into an independent company called AbbVie, it said on 2 January. Abbott, based in Abbott Park, Illinois, will retain its medical devices, diagnostics, nutritional products and generic pharmaceutical business. AbbVie, headquartered in North Chicago, Illinois, previously functioned as an Abbott subsidiary; it earned US$17.4 billion in 2011. The now-separate firm plans to expand the market for its blockbuster anti-inflammatory antibody Humira (adalimumab), and to push forward candidate drugs that include several hepatitis C antivirals.

Stem-cell transfer

Pioneering biotechnology company Geron is shedding its assets in human embryonic stem cells. The company, based in Menlo Park, California, announced its plans on 7 January. It will transfer cell lines, its early clinical programme in spinal-cord injury, and some 400 patents and patent applications to a subsidiary of BioTime, a company in Alameda, California, which is run by two former Geron heads. Geron funded initial academic work to isolate human embryonic stem cells in the 1990s, but current management said that the technology is a poor investment.

Fetal DNA sale

A firm that develops non-invasive tests for fetal DNA was sold to DNA-sequencing leader Illumina of San Diego, California, for US$350 million, the two companies announced on 7 January. Verinata Health, based in Redwood City, California, markets a test for chromosomal abnormalities, such as Down's syndrome, that involves sequencing fetal DNA circulating in a mother's blood. The company is embroiled in intellectual-property disputes with two other firms in a prenatal genetic-testing market potentially worth billions (see *Nature* **486,** 454; 2012). On the same day, Illumina's share price dropped by 7% after drug giant Roche said that it did not plan to buy the company.

Spill settlement

Drilling company Transocean, based in Vernier, Switzerland, reached a US$1.4-billion settlement with the US Department of Justice on 3 January for charges related to the 2010 Deepwater Horizon disaster, in which one of the company's oil platforms in the Gulf of Mexico exploded. The deal follows a $4.5-billion settlement last year by oil-and-gas giant BP (see *Nature* **491**, 501; 2012). It includes $150 million each for the National Academy of Sciences and the National Fish and Wildlife Foundation, to support environmental restoration and protection and health research in the region.

COMING UP

15–17 JANUARY
The nexus between water and energy comes under tighter focus at Abu Dhabi's annual World Future Energy Summit, which this year is held jointly with the inaugural International Water Summit.
go.nature.com/1b68r1

16–18 JANUARY
Experts on gravitational waves discuss frontiers in the field at a meeting in Seoul.
go.nature.com/jugqyh

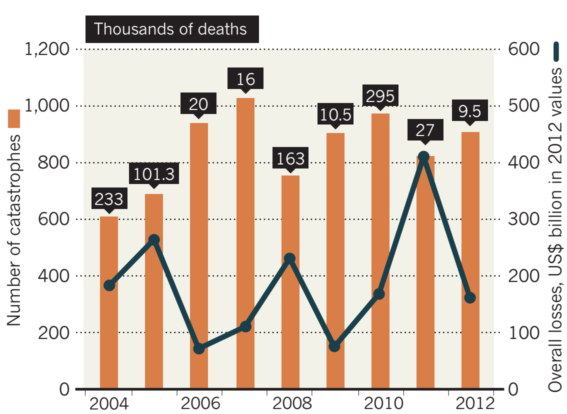
SOURCE: MUNICH RE

TREND WATCH

Around 9,500 people lost their lives last year in natural disasters — less than one-tenth of the ten-year average. Had it not been for Hurricane Sandy, material losses from storms, floods, droughts and earthquakes would also have been exceptionally low. Overall damage from natural disasters in 2012 was US$160 billion, less than half the 2011 figure and slightly below the ten-year average, according to the German reinsurance group Munich Re. Sandy alone accounted for an estimated $50 billion.

COUNTING CATASTROPHES

Deaths from natural disasters were relatively low in 2012, as few catastrophes occurred in developing countries.



NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

JOHN GUILLEMIN/BLOOMBERG VIA GETTY

The Belchatow power station in Poland is Europe’s largest coal-burning plant, but plans to capture carbon dioxide from it are in limbo.

CLIMATE

Europe’s untamed carbon

Funding and politics hobble CCS technology, seen as the best hope for cleaning up coal.

BY RICHARD VAN NOORDEN

With its carbon-trading market and tough emissions targets, Europe plays the part of responsible adult at climate-policy negotiations. But in a growing blemish on its low-carbon image, the region has fallen behind North America in the slow crawl to demonstrate systems for capturing greenhouse-gas emissions from power plants and industry — even as it increases its use of coal.

Announcements made just before Christmas underlined Europe’s troubles in launching large carbon capture and storage (CCS) projects. A European Commission fund set up two years ago in part to support CCS could not find a single scheme to finance, and instead gave €1.2 billion (US$1.6 billion) to renewable-energy projects. And in a mortifying side note, the International Energy Agency (IEA) noted that as the switch from coal to shale gas in the United States lowers the price of coal, Europe — where gas is expensive — is burning more of this dirtier fuel. “It is hugely embarrassing for Europe’s international standing on climate,” says Vivian Scott, who studies the role of CCS in climate policy at the University of Edinburgh, UK.

➲ NATURE.COM
To read more about CCS, see:
go.nature.com/enhtkr

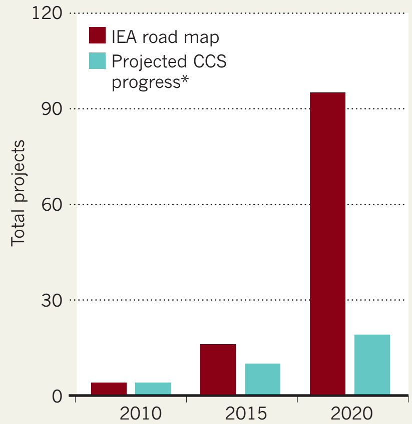
Admittedly, no country is finding CCS easy to fund. The technology to sieve carbon dioxide from exhaust gases has been demonstrated on a small scale, and four large projects have successfully stored the gas underground. ▶

But not one large carbon-capture system is operating at a power plant anywhere in the world. Fitting a large power plant with CCS would increase the price of its electricity by 50–100%, requiring hundreds of millions of dollars in subsidies.

Four years ago, the IEA set out a road map that called for building 100 CCS projects by 2020, together capturing around 150 million tonnes of carbon dioxide annually. Monitoring progress today, Scott says that the world will be lucky to see 20 completed in that time (see 'Carbon capture sputtering'). "The reality is that incentives in place for renewable energy are not in place for CCS. Government policies are not ambitious enough to drive large CCS projects," says Juho Lipponen, head of CCS at the IEA. Governments have pledged an estimated $20 billion for CCS demonstrations, but more like $100 billion would be needed to meet the IEA's original road map, he adds.

CARBON CAPTURE SPUTTERING

A 2009 road map from the International Energy Agency (IEA) foresaw carbon capture and storage (CCS) projects progressing at a much faster pace than is supported by current reality.



*Projections from Vivian Scott, Scottish Carbon Capture and Storage

SOURCE: IEA/VIVIAN SCOTT (SCCS)

EUROPEAN DELAYS

Yet the CCS industry in Europe has been hit particularly hard, by a combination of economic recession, political reluctance and Brussels bureaucracy. Europe has favoured wind and solar projects, with both direct grants and 'feed-in tariffs' that guarantee high prices for low-carbon electricity. Per unit of electricity, these subsidies have been as generous as those needed for CCS — but because the projects are smaller, individual payments are less daunting. CCS has had some direct grants but no feed-in tariffs; instead, plant operators were expected to benefit from Europe's emissions-trading scheme, under which polluting industries trade allowances to emit carbon dioxide. Utility companies that built CCS projects would be able to sell excess carbon credits, and the sale of 300 million credits would also finance a European CCS funding scheme (known as NER300).

> ***"Time will tell whether this technology has a market. For now it's important to see these plants up and running."***

But the price of credits collapsed to €6–7 per tonne last year, as the recession cut industrial emissions. Last year, a sale of 200 million credits raised only €1.5 billion, one-third of what planners had hoped for. "The collapse of the emissions-trading-scheme price is the biggest culprit for the lack of progress," says Howard Herzog, a sequestration researcher at the Massachusetts Institute of Technology in Cambridge.

Even so, the commission's fund had shortlisted ten CCS projects by July 2012, waiting for guarantees of co-funding from member states. But five of the six countries in which the plants were to be built — the United Kingdom, the Netherlands, Poland, Italy and Romania — "didn't step up to the plate," says Chris Davies, a British member of the European parliament who advocated for the CCS fund. Meanwhile, a plan to capture carbon emissions from a French steel plant became a casualty when owners ArcelorMittal, based in Luxembourg, pulled out of the scheme.

The missed opportunity has delayed CCS in Europe by two years, says Philippe Paelinck, who oversees CCS development at French engineering firm Alstom, based in Levallois-Perret outside Paris. Alstom runs carbon-capture technology at a number of pilot projects and had hoped to scale it up at three NER300 candidate projects. Nevertheless, the NER300 fund will award a second phase of grants, potentially this year; Paelinck thinks CCS projects in the United Kingdom, Romania and the Netherlands are well placed to gain some funding. Separate funding schemes, such as a £1-billion (US$1.6-billion) CCS competition in the United Kingdom — which is revamping its electricity market to allow tariff-like support for CCS — could also launch projects. "There's still activity for CCS in Europe: it's just proving to be a much harder sell than initially thought," says Lipponen.

A HELPING HAND FROM OIL

In the United States and Canada, CCS project developers have made faster headway by combining public subsidies with sales of the carbon dioxide they capture. Oil companies inject the gas — which sells for $20–40 per tonne — into the ground to help to loosen trapped oil. Two large North American projects are due to start capturing carbon in 2014: Canada's Can$1.24-billion (US$1.26-billion) Boundary Dam project in Saskatchewan, a 100-megawatt project that would rebuild one unit of an existing coal-fired plant; and the US$2.4-billion Texas Clean Energy Project near Odessa, a new 400-megawatt power plant that would gasify its coal before burning. A similar $2.4-billion plant in Kemper County, Mississippi, is also scheduled to start operation in 2014 but is currently facing a legal dispute over whether it can raise local electricity rates.

Funding CCS by striking deals with oil producers is not an option in oil-poor parts of Europe, however. And Stuart Haszeldine, a geologist at the University of Edinburgh, cautions that although selling carbon dioxide for oil extraction may help to get the industry going, it comes at the cost of the carbon emitted from the extra oil: he claims (in unpublished calculations) that once oil production is factored in, CCS funded in this way could actually cut emissions by only 10%.

Still, CCS advocates say that large-scale demonstrations of the technology are crucial because CCS offers the only way to clean up emissions from coal, which the IEA predicts will overtake oil to become the world's leading energy source by 2017. "Time will tell whether this technology has a market. For now it's important to see these plants up and running," says Julio Friedmann, head of the carbon-management programme at Lawrence Livermore National Laboratory in California. No country with long-term climate obligations could seriously meet them without CCS, he says. "Yet Europe is shutting down nuclear plants, building coal plants, and not keenly pursuing CCS. So the question is, is it serious about cutting emissions?" ■

MORE ONLINE

MICHAL BEDNAREK/THINKSTOCK

TOP STORY

● Ultracold quantum gas paves the way for negative-kelvin materials. go.nature.com/kpitlq

ON SOCIAL MEDIA

There are many ways to connect to Nature News:
● 'Like' us at Facebook.com/NatureNews
● Follow us on Twitter @NatureNews
● Friend us on Google+ at go.nature.com/vt51ek

MORE NEWS

● Martian meteorite matches findings from NASA rovers. go.nature.com/q8padi

DETLEV VAN RAVENSWAAY/SPL

OPTICS

Electron beams set nanostructures aglow

Physicists borrow an old tool from geology to probe the interior layers of metamaterials.

BY EUGENIE SAMUEL REICH

Put a piece of quartz under an electron microscope and it will shine an icy blue. First noticed in the 1960s, the phenomenon, called cathodoluminescence, gave geologists an easy way to identify quartz and other minerals in rock samples. But the light — emitted after a beam of electrons kicks a material's own electrons into a higher-energy state — is faint and diffuse, and that discouraged other scientists from harnessing it for fine-scale imaging.

Now a Dutch group has found a way to collect and focus a particularly faint and localized type of cathodoluminescence that had been previously ignored, turning the glow into a precise probe of a material's nanoscale structure. The researchers expect the technology to reach the market early this year, giving materials scientists a new tool for investigating the behaviour of light in the interiors of the complex nanostructures used in lasers, light-based circuits and solar cells.

"We have built a cathodoluminescent system that is unique in the world," says Albert Polman, a physicist at the FOM Institute for Atomic and Molecular Physics (AMOLF) in Amsterdam.

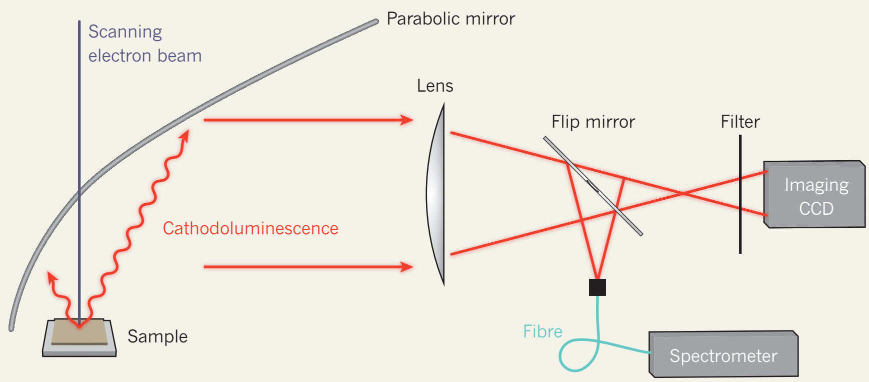
The technique combines the advantages of optical and electron-based imaging. An electron beam can in principle achieve a resolution of less than one nanometre, compared with hundreds of nanometres for a beam of light. But maps made by scattered or reflected electrons are not typically sensitive to the way light behaves in the sample. Cathodoluminescence, by contrast, can map the interaction of light and matter — but, because it is triggered by a narrow beam of electrons, it promises the same nanometre-scale resolution that those systems can achieve.

"This has opened the door to understanding how light couples to matter in a more fundamental way."

"This has opened the door to understanding how light couples to matter in a more fundamental way," says Harry Atwater, a physicist at the California Institute of Technology in Pasadena, who worked with Polman on cathodoluminescence in the past but is not involved with the work on the latest technique.

METAMATERIAL TV

A system of lenses and mirrors makes it possible to use cathodoluminescence, the principle of a cathode-ray-tube television, to map complex nanoscale structures within engineered 'metamaterials'.



SOURCE: ERNST JAN VESSEUR & TOON COENEN (AMOLF)

Key to the system is ultraprecise alignment of the sample. The device includes a carefully shaped parabolic mirror that collects photons as they emerge from a sample bombarded with electrons. It sends them through a system of lenses and a spectrometer to resolve the origin and energy of the light (see 'Metamaterial TV'). The device is sensitive enough to pick up a signal even from materials that are barely luminescent, such as metals. Just as in an old-fashioned cathode-ray-tube colour television, the electron beam scans the sample to build up an image line by line.

Already, the Dutch group, along with collaborators in the United States and Spain, has used the technique to tease out how certain nanostructures interact with light. In a recent paper studying a layered construct of silver, glass and silicon, they show that the phase velocity of visible light — the speed at which the peaks and troughs of the wave travel through the material — is so fast that it is effectively travelling in a vacuum, the explanation for the material's overall refractive index of zero (E. J. R. Vesseur *et al. Phys. Rev. Lett.* **110,** 013902; 2013).

The effect had been predicted in such layered structures, called metamaterials, but observing it required a higher-resolution map of light emission than earlier techniques could produce. The team has also mapped the distribution of light in the silicon nanodiscs that are used as a coating on solar cells to improve efficiency, and in the ultra-small cavities of photonic crystals — components of chip-based lasers and light-emitting diodes.

Lukas Novotny, an optical physicist at the Swiss Federal Institute of Technology in Zurich, says that cathodoluminescence could be a useful tool for improving the performance of light-emitting devices and solar cells, because the light-emission maps created with the technique reflect the local density of electromagnetic states, a quantity that determines how well light couples to matter and vice versa. "This information is key," he says.

The optics start-up company Delmic, based in Delft, the Netherlands, has licensed AMOLF's cathodoluminescence technique, and Polman says that the company will soon be selling the devices to materials researchers in universities for between US$100,000 and $200,000; later it may target the laser, semiconductor and solar-cell industries. He realizes that, by selling the system, he may create competitors for his own research. But he says that that will be more than made up for by the creation of a community of scientists using cathodoluminescence outside the ranks of geologists. "We want to motivate people to get out there and do this," he says. ■

SVEN TORFINN/PANOS PICTURES

Small-scale gold mining is a growing source of toxic mercury emissions.

ENVIRONMENT

Tough talk over mercury treaty

Nations debate how to share the costs of cutting emissions.

BY JANE QIU

Governments are on the verge of agreeing the first legally binding, global treaty to tackle mercury pollution. It aims to clean up the legacy of centuries of untrammelled emissions of the toxic metal, and to limit future contamination from sources as diverse as coal-fired power plants and gold mining.

Delegates from 128 countries are expected to meet next week in Geneva, Switzerland, for a fifth and final round of treaty negotiations. All agree that action is needed urgently to reduce mercury emissions, which pose risks to the environment and human health. But consensus on how to achieve that will not come easily. The current top emitters, in Asia, want to know why they should shoulder the burden of clean-up when much of the world's mercury pollution is due to the past economic growth of developed nations — an argument that parallels one of the main stumbling blocks to an international greenhouse-gas agreement.

Yet with mercury able to drift freely through air, soil, rivers and oceans, it is crucial that the negotiations deliver "a global treaty that is going to be implementable in all the countries", says Fernando Lugris, chairman of the Intergovernmental Negotiating Committee of the United Nations Environment Programme (UNEP), which is coordinating the treaty.

According to a draft version of UNEP's 2013 Global Mercury Report, about 6,500 tonnes of mercury was emitted into the air in 2010. Roughly 30% came from human activities, and a further 15% from natural sources such as volcanoes and erosion. The remainder was from the re-emission from soils, water and vegetation of mercury released into the environment decades ago. "Once emitted into the air, it's like a genie that has escaped the bottle," says David Streets, an environmental scientist at Argonne National Laboratory in Illinois. "It takes decades or even centuries to get it fixed."

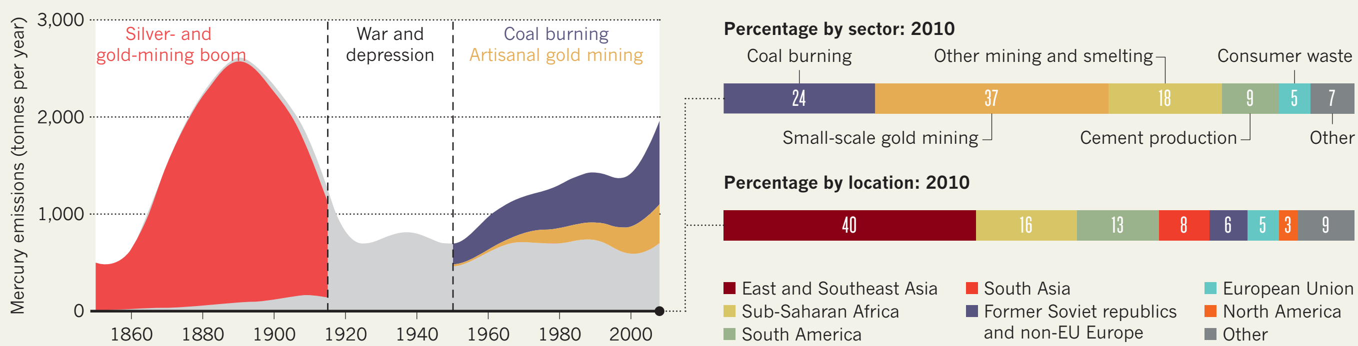
The UNEP report points out that mercury concentrations in the upper 100 metres of the oceans have doubled in the past century; top marine predators in the Arctic Ocean contain up to 12 times more mercury than in pre-industrial times, and contaminated seafood is a significant source of human exposure. The World Health Organization (WHO) rates mercury as one of the top ten chemicals of major public health concern — it can damage the brain and is a particular hazard for fetuses.

Streets estimates that human activity has released a grand total of about 350,000 tonnes of mercury, with roughly 40% of those emissions occurring before 1850 (see 'Quicksilver quantified'). Much of the rest came from silver and gold mining in the late nineteenth century, when mercury was used to extract precious metals. However, emissions have soared in recent decades, propelled by small-scale, or artisanal, gold mining in Africa and Latin America, and rapid industrialization in Asia (D. G. Streets *et al. Environ. Sci. Technol.* **45,** 10485–10491; 2011). China is now the biggest emitter, contributing 30% of global anthropogenic emissions.

Cleaning up industrial processes, for example by capturing mercury released from burning coal, involves costly technology. In the negotiations, "the biggest contention is who

QUICKSILVER QUANTIFIED

Anthropogenic mercury emissions are rising in Asia, but pollution more than a century old persists globally, raising questions about who should pay to clean up the metal.



SOURCE: D. G. STREETS *ET AL. ENVIRON. SCI. TECHNOL.* **45,** 10485–10491 (2011)/UNEP

should pay", says Ludovic Bernaudat, a mercury expert at the United Nations Industrial Development Organization in Vienna. "There are a lot of tensions right now."

Before the latest round of negotiations, in June 2012, the discussions aimed at an agreement for all countries to cap mercury emissions. But developing nations such as China and India — second only to China as an emitter — were adamant that this would be unfair unless developed nations helped with the cost and technologies. Common measures for controlling air pollution have the potential to reduce mercury emissions from coal plants by about 36%, "but to go further you'd need specific mercury-control technologies that can remove 90% of emissions, which are only available in developed countries", says Wang Shuxiao, an environment scientist at Tsinghua University in Beijing, who is part of the Chinese delegation.

Developed nations seem unlikely to commit to funding the transfer of such technologies. Negotiators may settle on an agreement that requires countries to set national targets that they can meet with the best mercury-control measures available to them, and to beef up monitoring programmes. UNEP predicts that such measures could reduce emissions in industrial regions by 25% by 2020, compared with an increase of up to 25% under a business-as-usual scenario.

The treaty also aims to limit emissions from artisanal gold mining, which is largely unregulated. Miners soak crushed ore in mercury to form an amalgam that leaves impurities behind; heating the amalgam frees the gold, but releases mercury into the air. "Most of them are unaware of the health hazards of mercury vapour and nobody wears a mask," says Nicola Pirrone, director of the Institute of Atmospheric Pollution Research in Rome. The treaty is likely to recommend that countries register and monitor mining, and will encourage technologies that capture mercury vapour or use jets of water and air to separate gold from ores (see *Nature* **486,** 306–307; 2012).

In late February, at the Global Ministerial Environment Forum in Nairobi, UNEP's governing council will debate the draft treaty that will emerge from next week's meeting. UNEP expects countries to ratify the treaty later this year. Even if the treaty does not set binding caps, "it should still build enough momentum for countries to commit to serious efforts to tackle the problem", says John Munthe, an environment-policy researcher at the Swedish Environmental Research Institute in Stockholm. "There are plenty of low-hanging fruits that could make a big difference in reducing global mercury emissions." ■

➲ NATURE.COM
Read more about pollution caused by gold mining:
go.nature.com/wlarqr

REGENERATIVE MEDICINE

Safety of induced stem cells gets a boost

Fears of immune response have been overestimated.

BY MONYA BAKER

A paper published in *Nature* today[1] could dispel a cloud over the hopes of turning a patient's own cells into perfectly matched replacement tissues.

Scientists first reported in 2007 that a person's cells could be reprogrammed to an embryo-like state, and so could form any type of cell in the body. Medical researchers immediately imagined using these 'induced pluripotent stem (iPS) cells' to create an endless supply of genetically matched replacement tissues to treat a range of diseases: fresh pancreatic tissue for diabetics, for example, or new nerve cells for people with Parkinson's.

The strategy also seemed to offer a way around the ethical complexities of using stem cells derived from human embryos. But then came the worries about possible side effects. Particularly bad news came from a 2011 study[2] showing that iPS cells provoked immune responses when injected into the mice from which they had been derived, casting doubt over one of the key advantages of the cells.

The latest *Nature* study[1] rejects that conclusion. Masumi Abe, a geneticist at the National Institute of Radiological Sciences in Chiba, Japan, and his team took iPS cells derived from mice and injected them back into the animals. For comparison, they injected other mice with embryonic stem (ES) cells. Yet unlike the 2011 study, which saw iPS cells perform worse than ES cells, the team found no differences between the immune responses of each group. The researchers also transplanted skin and bone-marrow cells derived from iPS or ES cells into mice and achieved similar success rates between the groups. The immune response of both sets of tissues is "indistinguishable", says Abe.

Konrad Hochedlinger, a stem-cell scientist at Massachusetts General Hospital in Boston, says that the result will probably "calm people down" about iPS cells. "It is definitely reassuring," he says.

The findings follow another positive study on iPS cells, published late last year[3], which found that the reprogramming process causes fewer mutations than previously thought. Flora Vaccarino, a neuroscientist at Yale University in New Haven, Connecticut, and her colleagues used high-resolution DNA analysis to compare the genomes of iPS cells and the adult cells from which they were derived. They found that most of the DNA mutations in the iPS cells did not arise during reprogramming but had been present in the parent cells.

Yang Xu, a stem-cell scientist at the University of California, San Diego, and co-author of the 2011 study[2], says that the new work does not dispel all concerns about the immune response provoked by iPS cells.

Xu points out that the skin and bone-marrow cells used in the latest study were not grown from iPS cells in culture, as they would be for clinical use. Instead, the researchers mixed iPS cells into early mouse embryos to make 'chimaeric' embryos. They then used skin and bone-marrow tissues that arose from iPS cells after the embryos grew into adult mice for their transplantation experiments. It is possible, says Xu, that the most immunogenic cells were rejected as the mice developed, which would explain why Abe and his colleagues observed a limited immune response. Transplanting tissues from chimaeric mice is "flawed", he says.

"The result will probably calm people down. It is definitely reassuring."

Producing chimaeric embryos is a standard technique for testing whether mouse iPS cells have been fully reprogrammed, says Jakub Tolar, a clinician at the University of Minnesota in Minneapolis, but he notes that differentiating cells in culture outside the body is much harder. Tolar, who hopes to use iPS cells to treat the childhood skin disease epidermolysis bullosa, adds that iPS-cell therapies will use human cells, which could behave quite differently from mouse cells. "It's helpful that they've done this, but it is absolutely different when you go to something that is cultured," he says.

Hochedlinger believes that iPS cells are just as promising for cell transplantation as ES cells, although many issues stand between the lab and the clinic. The differences between the two kinds of stem cell are minor compared with the differences in how individual cell lines grow and differentiate in culture, he says.

"Based on what we know at this time from mice," he says, "iPS cells are as good as ES cells, and should be as safe." ■

1. Araki, R. *et al. Nature* http://dx.doi.org/10.1038/nature11807 (2013).
2. Zhao, T., Zhang, Z.-N., Rong, Z. & Xu, Y. *Nature* **474,** 212–215 (2011).
3. Abyzov, A. *et al. Nature* **492,** 438–442 (2012).

should pay", says Ludovic Bernaudat, a mercury expert at the United Nations Industrial Development Organization in Vienna. "There are a lot of tensions right now."

Before the latest round of negotiations, in June 2012, the discussions aimed at an agreement for all countries to cap mercury emissions. But developing nations such as China and India — second only to China as an emitter — were adamant that this would be unfair unless developed nations helped with the cost and technologies. Common measures for controlling air pollution have the potential to reduce mercury emissions from coal plants by about 36%, "but to go further you'd need specific mercury-control technologies that can remove 90% of emissions, which are only available in developed countries", says Wang Shuxiao, an environment scientist at Tsinghua University in Beijing, who is part of the Chinese delegation.

Developed nations seem unlikely to commit to funding the transfer of such technologies. Negotiators may settle on an agreement that requires countries to set national targets that they can meet with the best mercury-control measures available to them, and to beef up monitoring programmes. UNEP predicts that such measures could reduce emissions in industrial regions by 25% by 2020, compared with an increase of up to 25% under a business-as-usual scenario.

The treaty also aims to limit emissions from artisanal gold mining, which is largely unregulated. Miners soak crushed ore in mercury to form an amalgam that leaves impurities behind; heating the amalgam frees the gold, but releases mercury into the air. "Most of them are unaware of the health hazards of mercury vapour and nobody wears a mask," says Nicola Pirrone, director of the Institute of Atmospheric Pollution Research in Rome. The treaty is likely to recommend that countries register and monitor mining, and will encourage technologies that capture mercury vapour or use jets of water and air to separate gold from ores (see *Nature* **486,** 306–307; 2012).

In late February, at the Global Ministerial Environment Forum in Nairobi, UNEP's governing council will debate the draft treaty that will emerge from next week's meeting. UNEP expects countries to ratify the treaty later this year. Even if the treaty does not set binding caps, "it should still build enough momentum for countries to commit to serious efforts to tackle the problem", says John Munthe, an environment-policy researcher at the Swedish Environmental Research Institute in Stockholm. "There are plenty of low-hanging fruits that could make a big difference in reducing global mercury emissions." ■

➲ **NATURE.COM**
Read more about pollution caused by gold mining:
go.nature.com/wlarqr

REGENERATIVE MEDICINE

Safety of induced stem cells gets a boost

Fears of immune response have been overestimated.

BY MONYA BAKER

A paper published in *Nature* today[1] could dispel a cloud over the hopes of turning a patient's own cells into perfectly matched replacement tissues.

Scientists first reported in 2007 that a person's cells could be reprogrammed to an embryo-like state, and so could form any type of cell in the body. Medical researchers immediately imagined using these 'induced pluripotent stem (iPS) cells' to create an endless supply of genetically matched replacement tissues to treat a range of diseases: fresh pancreatic tissue for diabetics, for example, or new nerve cells for people with Parkinson's.

The strategy also seemed to offer a way around the ethical complexities of using stem cells derived from human embryos. But then came the worries about possible side effects. Particularly bad news came from a 2011 study[2] showing that iPS cells provoked immune responses when injected into the mice from which they had been derived, casting doubt over one of the key advantages of the cells.

The latest *Nature* study[1] rejects that conclusion. Masumi Abe, a geneticist at the National Institute of Radiological Sciences in Chiba, Japan, and his team took iPS cells derived from mice and injected them back into the animals. For comparison, they injected other mice with embryonic stem (ES) cells. Yet unlike the 2011 study, which saw iPS cells perform worse than ES cells, the team found no differences between the immune responses of each group. The researchers also transplanted skin and bone-marrow cells derived from iPS or ES cells into mice and achieved similar success rates between the groups. The immune response of both sets of tissues is "indistinguishable", says Abe.

Konrad Hochedlinger, a stem-cell scientist at Massachusetts General Hospital in Boston, says that the result will probably "calm people down" about iPS cells. "It is definitely reassuring," he says.

The findings follow another positive study on iPS cells, published late last year[3], which found that the reprogramming process causes fewer mutations than previously thought. Flora Vaccarino, a neuroscientist at Yale University in New Haven, Connecticut, and her colleagues used high-resolution DNA analysis to compare the genomes of iPS cells and the adult cells from which they were derived. They found that most of the DNA mutations in the iPS cells did not arise during reprogramming but had been present in the parent cells.

Yang Xu, a stem-cell scientist at the University of California, San Diego, and co-author of the 2011 study[2], says that the new work does not dispel all concerns about the immune response provoked by iPS cells.

Xu points out that the skin and bone-marrow cells used in the latest study were not grown from iPS cells in culture, as they would be for clinical use. Instead, the researchers mixed iPS cells into early mouse embryos to make 'chimaeric' embryos. They then used skin and bone-marrow tissues that arose from iPS cells after the embryos grew into adult mice for their transplantation experiments. It is possible, says Xu, that the most immunogenic cells were rejected as the mice developed, which would explain why Abe and his colleagues observed a limited immune response. Transplanting tissues from chimaeric mice is "flawed", he says.

"The result will probably calm people down. It is definitely reassuring."

Producing chimaeric embryos is a standard technique for testing whether mouse iPS cells have been fully reprogrammed, says Jakub Tolar, a clinician at the University of Minnesota in Minneapolis, but he notes that differentiating cells in culture outside the body is much harder. Tolar, who hopes to use iPS cells to treat the childhood skin disease epidermolysis bullosa, adds that iPS-cell therapies will use human cells, which could behave quite differently from mouse cells. "It's helpful that they've done this, but it is absolutely different when you go to something that is cultured," he says.

Hochedlinger believes that iPS cells are just as promising for cell transplantation as ES cells, although many issues stand between the lab and the clinic. The differences between the two kinds of stem cell are minor compared with the differences in how individual cell lines grow and differentiate in culture, he says.

"Based on what we know at this time from mice," he says, "iPS cells are as good as ES cells, and should be as safe." ■

1. Araki, R. *et al. Nature* http://dx.doi.org/10.1038/nature11807 (2013).
2. Zhao, T., Zhang, Z.-N., Rong, Z. & Xu, Y. *Nature* **474,** 212–215 (2011).
3. Abyzov, A. *et al. Nature* **492,** 438–442 (2012).

ETHICS

'Rehab' helps errant researchers return to the lab

As cases of misconduct rise, ethicists test an approach for reforming offenders.

BY DANIEL CRESSEY

On 8 January, four health scientists accused of research misconduct are expected to file into a room at Saint Louis University in Missouri to conduct an experiment: can they be rehabilitated? Using pseudonyms if they wish, the four will discuss what they did wrong, why they did it and how they might stop it from happening again — all in an effort to see if they can re-enter the scientific mainstream.

With the rapid growth of misconduct cases, some scientists are worried that preventative training in research ethics might not be enough. Nor will it be possible simply to dismiss all violators from science. Scientific rehabilitation, they say, will have to become a necessary tool for research-integrity offices.

"Sometimes these are very talented researchers," says James DuBois, an ethicist at Saint Louis University, who leads the rehab programme, called RePAIR (Restoring Professionalism and Integrity in Research). "We believe that if we can equip them with certain skills, they can return to the field as very productive individuals."

DuBois could have plenty of customers on the way. The US Office of Research Integrity (ORI) in Rockville, Maryland, received 419 allegations of misconduct at institutions in 2012 — nearly double the number in 2011, says David Wright, director of the ORI. Wright puts much of this down to software that has made it easier for watchdogs to detect possible plagiarism or image manipulation in publications. "We now get allegations from all over the world," he says.

As the pursuit of misconduct increases, there has been less focus on how to deal with guilty scientists. Lauran Qualkenbush, director of the research-integrity office at Northwestern University in Chicago, says that there is a gap between harsh penalties for misconduct — such as bans from receiving government funding — and lighter punishments such as taking a short online ethics course and getting a letter of reprimand. "It's not always black and white," she says. "These are people we feel could be valuable members of our faculty and community."

DuBois aims to fill that gap. Although remedial courses have been available for physicians for more than a decade — with many returning to medicine to forge successful careers — DuBois says that RePAIR is the first such programme for researchers. Developed

BRIAN CASSIDY/SAINT LOUIS BUSINESS JOURNAL

James Dubois hopes that his programme will let wayward researchers back into the scientific fold.

with a US$500,000 grant from the National Institutes of Health, the three-day course costs $3,000 per participant, or half that if host institutions sign up as partners for $1,500 a year.

Day one of the course focuses on 'self-serving biases' — in which people take credit for success but blame external factors for failures — and on getting participants to examine the factors that have led to them breach the accepted rules of research. The second day gets participants to talk about what they did wrong, and focuses on skills to promote ethical decision-making, such as what they could have done differently in their own research. On the final day, professional management plans are drawn up for each participant to try to prevent them falling back into bad habits, with follow-up interviews taking place in the weeks after the course. If participants complete all of these activities to the satisfaction of the RePAIR team, a certificate of completion is sent to their institution.

Last year, DuBois conducted a survey of 129 US institutions and found that most were investigating between three and five cases of wrongdoing per year. He says that, so far, around six universities, including Northwestern, have signed up as partners in the programme.

Some ethicists are unsure how effective such rehab will be. Nicholas Steneck, an ethicist at the University of Michigan in Ann Arbor, is broadly supportive of the goals of RePAIR, but says that he wonders if the money might be better spent on measures that prevent misconduct.

Although the ORI is not endorsing RePAIR specifically, Wright says that his previous job as a research-integrity officer at Michigan State University in East Lansing has made him well aware of the need for such a programme, especially for researchers who have committed low-level misconduct, such as failing to disclose a conflict of interest. "What do you do with them to make them whole as members of the research community, so they're not just pariahs in their community forever? My view is if you're not going to fire somebody, there ought to be a way for them to re-establish themselves."

Wright says it is too early to say if DuBois and colleagues' programme will work, but thinks that it is "a worthy effort and I salute them".

Scrutiny of misconduct is also growing rapidly in Europe. According to Steneck, who consults for Epigeum, a London-based company that provides training on research integrity, that is because the definition of misconduct is broadening beyond falsification, fabrication and plagiarism. 'Grey area' cases such as self-plagiarism and duplicate publication are increasingly attracting attention, he says. However, the extent of misconduct is harder to establish in Europe than in the United States, because the overarching European Union research-funding system — as well as those of individual countries like the United Kingdom — lack an official investigatory body similar to the ORI. This means that responsibility for dealing with misconduct can fall solely on individual institutions.

A session of the RePAIR programme in May has already been advertised, and DuBois eventually hopes to run the course for up to ten people at a time.

"A lot of people will say to me, 'These are bad apples — you're not going to reach them,'" he says. "My sense is that may be true of 10%, 15%, maybe even 20%, but I do believe that we can make a difference in the lives of the vast majority." ■

CORRECTION

The News Feature '*Nature*'s 10' (*Nature* **492,** 335–343; 2012) gave the wrong age for Jun Wang. He is 36, not 34.

Brian Butterworth is on a crusade to understand the number deficit called dyscalculia — and to help those who have it.

NUMBER GAMES

BY EWEN CALLAWAY

In the mid-1980s, Paul Moorcraft, then a war correspondent, journeyed with a film crew into Afghanistan to produce a documentary about the fifth anniversary of the Soviet invasion. The trip took them behind Soviet lines. "We were attacked every fucking day by the Russians," says the colourful Welshman. But the real trouble started later, when Moorcraft tried to tally his expenses, such as horses and local garb for his crew. Even with a calculator, the simple sums took him ten times longer than they should have. "It was an absolute nightmare. I spent days and days and days." When he finally sent the bill to an accountant, he had not realized that after adding a zero he was claiming millions of pounds for a trip that had cost a couple of hundred thousand. "He knew I was an honest guy and assumed that it was just a typo."

Such mistakes were part of a lifelong pattern for Moorcraft, now director of the Centre for Foreign Policy Analysis in London and the author of more than a dozen books. He hasn't changed his phone number or PIN in years for fear that he would never remember new ones, and when working for Britain's Ministry of Defence he put subordinates in charge of remembering safe codes. In 2003, a mistaken phone number — one of hundreds before it — lost him a girlfriend who was convinced he was out gallivanting. That finally convinced him to seek an explanation.

At the suggestion of a friend who teaches children with learning disabilities, Moorcraft contacted Brian Butterworth, a cognitive neuroscientist at University College London who studies numerical cognition. After conducting some tests, Butterworth concluded that Moorcraft was "a disaster at arithmetic" and diagnosed him with dyscalculia, a little-known learning disability sometimes called number blindness and likened to dyslexia for maths. Researchers estimate that as much as 7% of the population has dyscalculia, which is marked by severe difficulties in dealing with numbers despite otherwise normal (or, in Moorcraft's case, probably well above normal) intelligence.

That combination has attracted neuroscientists such as Butterworth, who believe that the disorder illuminates the inner workings of the brain's number sense — the ability to understand and manipulate quantities. This sense is every bit as innate as vision or hearing, yet scientists disagree over its cognitive and neural basis, a debate that dyscalculics may help to settle.

For Butterworth, scientific curiosity eventually gave way to advocacy. "I thought, it's not enough to just try to identify the cause," he says. In the past decade, he has crusaded to get dyscalculia recognized — by parents, teachers, politicians and anyone who will listen. And he is using his scientific insights into the condition to help dyscalculic children. "What's the point of telling someone they have dyscalculia if you can't help them?" he says.

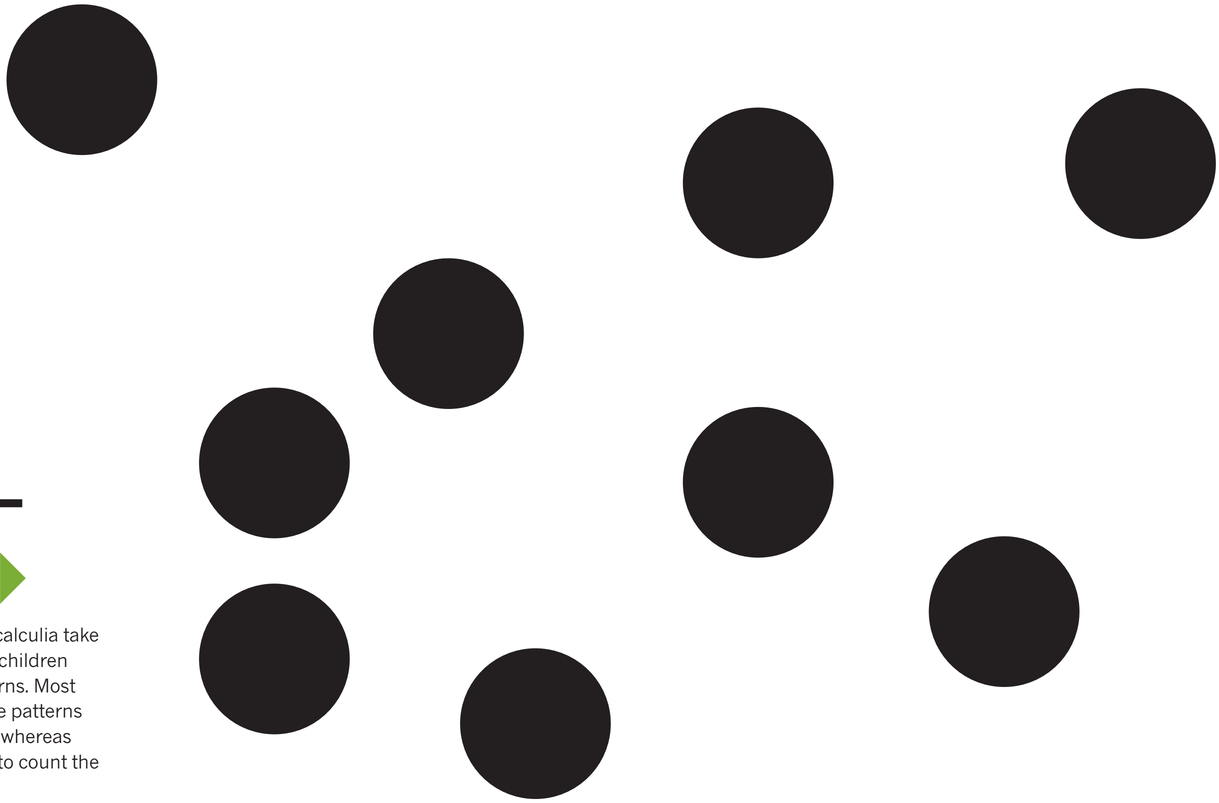
FINDING THE NUMBER

Christopher, a chatty nine-year-old in a rumpled blue sweatshirt and white polo shirt, sits beside Patricia Babtie, a teacher who specializes in dyscalculia and tutors children across Greater London. On a sturdy-looking laptop, Christopher (not his real name) is navigating Number Sense, a suite of educational computer games designed by Butterworth and his colleague Diana Laurillard at the Institute of Education in London.

By developing treatments for dyscalculia, Butterworth hoped to test competing theories about the cognitive basis of numeracy. If, as he believes, dyscalculia is at heart a deficiency of basic number sense and not of memory, attention or language, as others have proposed, then nurturing the roots of number sense should help dyscalculics such as Christopher. "It may be the case that what these kids need is just much more practice than the rest of us," Butterworth says. Christopher's school is one of several in London working with the software, and students in Cuba, Singapore and elsewhere will also soon start using it.

➲ NATURE.COM
For more on this story listen to the *Nature* podcast:
go.nature.com/eho7ig

Christopher starts with a game involving a number line — a spatial representation that scientists believe is key to number sense. "What is the number that is right in the middle between 200 and 800? Do you



HOW FAST CAN YOU COUNT THESE DOTS?

Children with dyscalculia take longer than other children to count dot patterns. Most instantly recognize patterns of up to four dots, whereas dyscalculics tend to count the dots one by one.

know it?" Babtie asks. Christopher shrugs. "Think of any number that is bigger than 200 and smaller than 800 and put it in this box. It could be 201," she says. He enters 200, and Babtie reminds him that the number needs to be greater than 200. He selects 210, probably mistaking it for 201. A classic sign of dyscalculia is difficulty in grasping the place-value system, according to Babtie. "That will do fine," she says. A soft computer voice tells Christopher to "find the number and click it". The game involves zooming in and zooming out to rescale the number line, and Christopher talks through each move — a strategy that Babtie encourages — but it takes him more than a minute to locate 210. His classmates, meanwhile, are learning to multiply two-digit numbers.

Some children at Christopher's school have more profound numeracy problems. One nine-year-old classmate says that she doesn't know if 50 is greater or less than 100; another the same age confuses four dots for five and routinely tots up small sums on his fingers, a common strategy for dyscalculics.

"OK, time to stop. We'll do some more of this another day," Babtie says to Christopher, after 20 frustrating minutes. It is clear he would rather be back in class than here in this room practising a skill his classmates learned years ago.

HOW MANY ARE THERE?

Butterworth, now 69, straddles the academic and public spheres. A fellow of the British Academy, the United Kingdom's national body for the humanities and social sciences, he made his name probing obscure speech and language disorders and has appeared in the British media for many years. In a *Sunday Times* article in 1984, for example, Butterworth claimed that the speech patterns of former US president Ronald Reagan indicated Alzheimer's disease. Reagan was diagnosed with the condition a decade later.

In the late 1980s Butterworth studied a stroke patient who would change the course of his professional life. The woman, a 59-year-old former hotel manager from Italy, fared about average on verbal IQ tests and had a good memory, but when Butterworth's Italian colleagues asked her to count, she would start, "*uno, due, tre, quattro*," and then stall. "*Miei matematica finisce alle quattro*" — my mathematics stops at four — the woman, known as CG, would tell them[1].

Neurologists had presented case studies of 'acalculic' patients such as CG from the early twentieth century onwards, if not before, but "people hadn't thought a lot about the specific brain areas involved in calculation", says Butterworth. Brain scans of CG revealed a lesion in the parietal lobe, a part of the brain just above the ears. Later, Butterworth found another patient with the opposite pattern of disability: neurodegeneration had robbed him of speech, language and much of his knowledge, save for the ability to do intricate calculations. Butterworth grew more certain that numerical abilities relied on specialized brain networks, and not only on those supporting general intelligence, as many scientists believed at the time.

Genetics and the vagaries of brain development disrupted these networks in dyscalculics, Butterworth proposed. And Moorcraft was one of Butterworth's most revelatory subjects because of the great disparity in his abilities in different domains. Butterworth and his colleagues also tested 31 eight- and nine-year-old children who were near the bottom of their class in mathematics but did well enough in other subjects. Compared with normal children and those with dyslexia, the dyscalculic children struggled on almost every numerical task, yet were average on tests of reading comprehension, memory and IQ[1].

The study confirmed for Butterworth that developmental forms of dyscalculia are the result of basic problems in comprehending numbers and not in other cognitive faculties. But determining exactly what those problems are would prove challenging.

Like nearly all human cognitive abilities, number sense is evolutionarily ancient — tens if not hundreds of millions of years old. Studies of chimpanzees, monkeys, newborn chicks, salamanders and even honeybees point to two parallel systems for representing quantities. One, called the approximate number sense, distinguishes larger quantities from smaller ones, be they dots flashing on a screen or fruits in a tree. Studies on monkeys reveal that

PETE REPKA

Brian Butterworth hopes that his number games will help dyscalculic children — and open a window on how the brain processes numbers.

certain neurons in a specific fold of the parietal lobe fire more vigorously in response to increasingly higher numbers[2]. A second ancient number system allows humans and many other animals to instantly and precisely recognize small quantities, up to four. Primate studies show that individual neurons within the same fold, called the intraparietal sulcus, seem tuned to particular quantities, such that when a monkey is performing a task that involves numbers, one neuron will fire for the number 1, a different one will fire for 2 and so on[3].

People who are poor at distinguishing approximate quantities do badly in maths, suggesting that the approximate-number system is crucial[4]. And some work shows that dyscalculics are poor at recognizing small numbers, suggesting that this ability is also fundamental to numeracy[5]. Moreover, scans of people with dyscalculia suggest that their intraparietal sulci are less active when processing numbers[6] and less connected with the rest of the brain[7] compared with numerate children and adults.

Yet Butterworth views such results as consequences, not causes, of the poor numerical abilities that characterize dyscalculia. He argues that another cognitive capacity is even more fundamental to number sense. He calls this 'numerosity coding': the understanding that things have a precise quantity associated with them, and that adding or taking things away alters that quantity.

But Stanislas Dehaene, a cognitive neuroscientist who studies numerical cognition at INSERM, France's national institute for research on medicine and health, near Paris, sees number sense as being supported by a broader set of cognitive features. Approximation and a sense of small numbers, while critical, are not enough for humans to precisely grasp large numbers, he says. Language, he argues, empowers humans to integrate the two number systems

WHICH NUMBER IS LARGER?

These tests are used to determine whether a person's problems with numbers are due to dyscalculia or to other cognitive deficits.

3 8

WHICH IS TALLER?

People with dyscalculia answer this question just as quickly and accurately as people without a learning disability.

7 9

WHICH IS GREATER?

Dyscalculics take longer and are less accurate at answering this question. They have even more trouble when the difference between the two numbers is small.

— giving them the ability to intuitively distinguish, say, 11,437 from 11,436. Butterworth's concept of numerosity coding may be an important part of number sense, says Dehaene, but there is still much to learn about it — for instance, whether it is present in other animals or in children from a very early age.

One of Butterworth's favourite papers is titled 'Six does not just mean a lot: preschoolers see number words as specific'[8]. In it, the developmental psychologist Barbara Sarnecka, now at the University of California, Irvine, and Susan Gelman, at the University of Michigan in Ann Arbor, showed that young children who could not yet count past two nonetheless understood that adding pennies to a bowl containing six somehow altered its number, even if the children couldn't say exactly how. If numerosity coding is fundamental, it predicts that dyscalculics such as Moorcraft or

Christopher struggle to enumerate and manipulate all numbers, large and small. Butterworth hopes that, by honing this ability, the Number Sense games will help support his research ideas.

Three months on, Christopher seems to be faring better at the number-line game, going so quickly that Babtie asks him to slow down and explain his reasoning for each move. Babtie says that dyscalculic children tend to learn much more quickly when they talk through what they do. She also believes that Christopher's maths anxiety, a near-universal trait of child and adult dyscalculics, is fading.

He moves on to a Tetris-like game called Numberbonds, in which bars of different lengths fall down the screen, and he is asked to select a block of the correct size to fill out a row. This emphasizes spatial relationships, which some dyscalculics also struggle with. The blocks move too quickly at first, frustrating Christopher, but he soon gets the hang of it, and when Babtie suggests he stop for the day, he begs for ten more minutes.

The Number Sense games, including a snazzy-looking iPhone version of Numberbonds, are intended to nurture the abilities that, Butterworth contends, are the root of numerical cognition and the core deficit of dyscalculia — manipulating precise quantities. In a game called Dots to Track, for example, children must ascribe an Arabic numeral to a pattern of dots, similar to those on dice. When they enter the wrong value — and they often do — the game asks the children to add or remove dots to achieve the correct answer.

As the summer holidays approach, Babtie is worried that Christopher and the other students she has been working with won't practise the games at home, returning in the autumn the worse for it. But in early October when school is back, Christopher announces that he will challenge himself with a number line that stretches from 950 to 9,000, "if you'll allow me", he adds. At first he flounders, but quickly starts to understand the game and locates a string of four-digit numbers, beaming with each correct response.

Other students are improving more slowly, but it is not easy to say why. Dyslexia, attention deficit hyperactivity disorder and autism spectrum disorder are common among dyscalculics, and it can be difficult to untangle these problems, says Babtie. The nine-year-old who counted on his fingers nine months ago can now deal with numbers below 6, but still struggles to distinguish 9 from 10. Yet with the right practice and attention from teachers and parents, dyscalculic children can thrive, says Babtie, who emphasizes that computer games are a supplement, not a replacement, for one-on-one tutoring.

Butterworth knows that it will take a controlled evaluation of Number Sense before he can say if the game genuinely improves numeracy in dyscalculic children. Small studies of other computer-based interventions hint that they might help. Dehaene reported in 2009 that Number Race, a game his group developed, modestly improved the ability of 15 dyscalculic kindergarten children to discern the larger of two numbers, but that it had no effect on their arithmetic or counting[9]. Meanwhile, a Swiss team reported in 2011 that a game that involves placing a spaceship on a number line helped eight- to ten-year-old dyscalculics with arithmetic. The researchers also studied the children in an fMRI scanner during a task that involved arranging numbers. They found that one month after training, the children showed increased activation in the intraparietal sulcus and reduced neural activation elsewhere in the parietal lobes — a hint that their improvements in arithmetic were related to changes involving brain areas that respond to number[10].

Butterworth hopes to monitor the brains of students such as Christopher as they practise Number Sense, to see if their parietal lobes are indeed changing. But he has been turned down by every funding source he has applied to. Although dyscalculia, like other learning disabilities, takes a toll on productivity (one report estimated that low numeracy costs the United Kingdom £2.4 billion (US$4 billion) per year, mostly in lost wages) it doesn't attract much attention or money. In the United States, for example, the National Institutes of Health spent $2 million studying dyscalculia between 2000 and 2011, compared with more than $107 million on dyslexia.

"EVERY DAY THEY GO TO SCHOOL. EVERY DAY THERE'S A MATHS CLASS. EVERY DAY THEY'RE SHOWN UP TO BE INCOMPETENT."

Butterworth's team now has tentative plans to evaluate its software with researchers at the Cuban Neurosciences Center and the University of Pedagogical Sciences in Havana next year, and the group is also placing the game in other countries, including China and Singapore. "The Cubans, curiously, are putting money into this, even though they've got very little," Butterworth says, commending the strength of the country's education system.

Although an emeritus professor, technically retired, Butterworth continues to research the neurodevelopmental roots of number sense, recently showing that guppies[11], like humans, possess approximate and precise number systems, and that dyscalculic adults have no more trouble telling the time than numerate people[12].

He hopes that Number Sense — if it can improve dyscalculia — will help him in the academic debate over the cognitive basis of numeracy. But Dehaene, probably his most fervent opponent in that debate, isn't counting on classroom computer games to resolve it. His Number Race game and its successor, Number Catcher, incorporate a multitude of numerical skills, so even if the game works, it won't address the theoretical differences about which skills are most essential to number sense or most compromised in dyscalculics. "I quickly realized that the interest of the children was to have a fun game full of ideas and variety, and that was not very compatible with an analytic approach," he says.

Butterworth, too, says that he is ultimately more motivated by helping children. In the course of his studies, he was struck that children "were very, very distressed by being bad at maths. So every day they would go to school, every day there's a maths class, every day they're shown up to be incompetent in a way other kids in their class are not", he says.

Moorcraft can commiserate. When he occasionally meets dyscalculic children, he tells them that he, too, counts with his fingers under the table, that they have nothing to be embarrassed about and that, with the practice that he never got, they can get up to speed.

Moorcraft is also completing a book on dyscalculia with one of Butterworth's postdocs. "I have written an introduction," he says. "I just hope the chapters are in the right order." ■

Ewen Callaway *writes for* Nature *from London.*

1. Ciplotti, L., Butterworth, B. & Denes, G. *Brain* **114,** 2619–2637 (1991).
2. Roitman, J. D., Brannon, E. M. & Platt, M. L. *PLoS Biol.* **5,** e208 (2007).
3. Sawamura, H., Shima, K. & Tanji, J. *Nature* **415,** 918–922 (2002).
4. Halberda, J., Mazzocco, M. M. & Feigenson, L. *Nature* **455,** 665–668 (2008).
5. Koontz, K. & Berch, D. *Mathematical Cognition* **2,** 1–23 (1996).
6. Price, G. R. *et al. Curr. Biol.* **17,** R1042–R1043 (2007).
7. Rykhlevskaia, E. *et al. Front. Hum. Neurosci.* **3,** 51 (2009).
8. Sarnecka, B. W. & Gelman, S. A. *Cognition* **92,** 329–352 (2004).
9. Räsänen, P. *et al. Cogn. Dev.* **24,** 450–472 (2009).
10. Kucian, K. *et al. NeuroImage* **57,** 782–795 (2011).
11. Agrillo, C. *et al. PLoS ONE* **7,** e31923 (2012).
12. Cappelletti, M. *et al. Front. Psychol.* **2,** 364 (2011).

IT COULD HAPPEN ONE NIGHT

CATASTROPHES FROM THE PAST WILL STRIKE AGAIN — WE JUST DO NOT KNOW WHEN.

BY NICOLA JONES

One hundred thousand years ago, a massive chunk of the Mauna Loa volcano cracked away from Hawaii and slid into the sea, launching a wave that rose as high as the Eiffel tower up the slopes of a nearby island. That mega-tsunami was not an isolated incident: the past 40,000 years have seen at least ten gigantic landslides of more than 100 cubic kilometres in the North Atlantic ocean alone, each capable of producing waves tens to hundreds of metres high. Another is bound to happen sometime — although whether it will strike tomorrow or 10,000 years from now is anyone's guess.

This week, the World Economic Forum published its 2013 global risks report, which includes a section, produced in collaboration with *Nature*, on X factors: low-probability, high-impact risks resulting mainly from human activity (see go.nature.com/outhzr). But the natural world holds unpredictable threats as well. The geologic record is peppered with evidence of rare, monstrous disasters, ranging from asteroid impacts to supervolcanoes to γ-ray bursts. *Nature* looks into some of the life-shattering events that Earth and the broader Universe could throw our way.

DEATH BY VOLCANO

Earth is now in the middle of a flare-up of supervolcanic activity[1]. Over the past 13.5 million years, no fewer than 19 giant eruptions have each spewed more than 1,000 cubic kilometres of rock — enough to coat an entire continent in a few centimetres of ash and push the planet into 'nuclear winter'. One of the most recent such eruptions, of Toba in Indonesia 74,000 years ago, was such a catastrophic event that some scientists have blamed it for starting the last ice age and slashing the human population to about 10,000 people. One estimate[1] suggests that there is a 1% chance of a super-eruption in the next 460–7,200 years.

The four youngest, most active supervolcanic systems in the world are Toba, Campi Flegrei in Italy, Yellowstone in the northwestern United States and Taupo in New Zealand. All four systems are being monitored for groundswell and seismic swarms — clusters of small earthquakes that can signal moving magma — and all occasionally show these warning signs. But no one knows whether the result of each flare-up will be a small squirt of steam or — much more hazardous — a mega-eruption of lava. "If something were brewing, we would get warning hours, days and months ahead," says Shan de Silva, a volcanologist at Oregon State University in Corvallis. "But how big it's going to be, we don't have a handle on."

To help answer these questions, scientists are now drilling into the heart of one of the top contenders for the next blow-up: the Campi Flegrei caldera, a crater that is 13 kilometres wide and includes the city of Naples. Since 1969, the ground at Campi Flegrei has bulged upwards by as much as 3.5 metres, and researchers are eager to find out whether the culprit is underground steam or a pool of magma. Previous bouts of volcanic activity in the caldera came after the ground surface had swelled up by several metres or more[2], and researchers think that major activity could occur within the next few decades or centuries. To investigate the risk, scientists at Campi Flegrei plan to drill more than 3 kilometres into the crater, despite concerns from some researchers that the drilling could trigger earthquakes or an explosion.

ANDY POTTS

One goal is to look at the magma pool beneath the crater: the shallower and more molten it is, the greater the chances of a super-eruption. Characterizing such pools through seismic studies is hard, and the range of error is huge. "We really are groping in the dark," says de Silva. Scientists estimate that 10–30% of the magma under Yellowstone, for example, is liquid — shy of the 50% thought to be needed for super-eruption. But pockets of molten magma in the chamber could still cause eruptions several-fold larger than the 1980 blast from Mount St Helens in Washington state, warns Jacob Lowenstern, head of the Yellowstone Volcano Observatory for the US Geological Survey in Menlo Park, California.

The effort to drill into Campi Flegrei and measure features such as temperature and rock permeability should help researchers to interpret seismic-imaging studies of magma pools, says Lowenstern. "If we want to be able to successfully image Earth, we occasionally need to make a few strategic incisions into the patient," he says. As for the dangers of drilling, Lowenstern is convinced that the project will have minimal impact. "It's like a pinprick on an elephant," he says. The Campi Flegrei team finished an initial 500-metre test well in December 2012 without incident. And seismologists safely drilled a hole of similar size into the Long Valley caldera in California — a supervolcano site that erupted 760,000 years ago and holds the same killer potential as Yellowstone.

Until more is learned about these systems, societies must accept that the threat of a super-eruption is real, yet remote. Lowenstern says that although the chances of one happening this year are tiny, "it is theoretically possible".

DEATH BY FUNGUS

Although viruses and bacteria grab more attention, fungi are the planet's biggest killers. Of all the pathogens being tracked, fungi have caused more than 70% of the recorded global and regional extinctions[3], and now threaten amphibians, bats and bees. The Irish potato famine in the 1840s showed just how devastating such pathogens can be. *Phytophthora infestans* (an organism similar to, and often grouped with, fungi) wiped out as much as three-quarters of the potato crop in Ireland and led to the death of one million people.

Potato blight is still a threat: 13_A2, a highly aggressive strain of *P. infestans*, is now rampant in Europe and North Africa. Across the globe, *Phytophthora* causes some US$6.7 billion in annual damages, according to a 2009 estimate[4]. Sarah Gurr, a plant pathologist at the University of Oxford, UK, estimates that the worst theoretical potato infestation would deprive 1.3 billion people of food each year. Other major staple crops face similar threats, such as rice blast (*Magnaporthe oryzae*), corn smut (*Ustilago maydis*), soya bean rust (*Phakopsora pachyrhizi*) and wheat stem rust (*Puccinia graminis*). The stem-rust superstrain Ug99 has in recent years slashed yields in parts of Africa by as much as 80%.

If all five crop staples were hit with fungal outbreaks at the same time, more than 60% of the world's population could go hungry, says Gurr. "That's apocalyptic", but unlikely, she says — "more of a James Bond movie". David Hughes, a zoologist at Pennsylvania State University in University Park, adds that terrorists could use fungi to wreak havoc by targeting economically important crops. In the 1980s, for example, a possibly deliberate infection wiped out cacao crops in northern Brazil, changing the country's demographics and ecology as people moved from

➲ NATURE.COM
Hear more about natural catastrophes on *Nature*'s podcast: go.nature.com/7kxzjw

unproductive farms to the cities and cleared more rainforest. "If you wanted to destabilize the world, you could easily introduce rubber blight into southeast Asia," he says, which would trigger a chain reaction of economic and political effects.

Modern agriculture has exacerbated societies' vulnerability by encouraging farmers to plant the same strains of high-yield crops, limiting the variety of resistance genes among the plants, says Gurr. "We've skewed the arms race in favour of the pathogen," she says. "That's why we're on the brink of disaster."

Researchers estimate that there are 1.5 million to 5 million species of fungi in the world, but only 100,000 have been identified. Reports of new types of fungal infection in plants and animals have risen nearly tenfold since 1995 (ref. 3). Gurr suggests that climate change might be a culprit.

Humans have cause for concern as well. In the past decade, a tropical fungus called *Cryptococcus gattii* has adapted to thrive in cooler climes and invaded the forests of North America's Pacific Northwest. By 2010, it had infected some 280 people, dozens of whom died. Although fungi are not spread as easily from person to person as viruses, for example, and anti-fungal agents can effectively tackle most infections, there are still reasons to worry. Fungi continue to evolve, and once they are established in an ecosystem, they can be almost impossible to wipe out.

Given these trends, experts say that fungi have not received enough attention from researchers and governments. "I'd be very surprised if an abrupt fungal infection killed a large swathe of people. But it's not impossible," says Matthew Fisher, an emerging-disease researcher at Imperial College London. "Complacency is not a recommended course of action."

A GIANT SOLAR FLARE COULD KILL HUNDREDS OF MILLIONS AND SET US BACK 150 YEARS.

DEATH FROM ABOVE

The heavens hold plenty of threats. The Sun occasionally launches outsize solar flares, which fry electricity grids by generating intense currents in wires. The most recent solar megastorm, in 1859, sparked fires in telegraph offices; today, a similarly sized storm would knock out satellites and shut down power grids for months or longer. That could cause trillions of dollars in economic damage.

A solar flare some 20 times larger than that may have hit Earth in 774, according to Adrian Melott, a cosmologist at the University of Kansas in Lawrence, and Brian Thomas, an astrophysicist at Washburn University in Topeka, Kansas. "That's not an extinction event," says Melott, "but for a technological civilization, it could kill hundreds of millions of people and set us back 150 years." Fortunately, there are ways to mitigate this worst-case scenario should it occur: engineers can protect the grid with fail-safes or by turning off the power in the face of an incoming blast.

Next up the scale of disaster magnitude is a large comet or asteroid strike. Sixty-five million years ago, an asteroid 10 kilometres wide hit Earth and triggered the end-Cretaceous mass extinction; 2-kilometre rocks, thought to be capable of causing extinctions on a smaller scale, smack the planet once or twice every million years. Astronomers are hard at work tallying and tracking asteroids in Earth's vicinity, and scientists are investigating ways to divert any real threats that might materialize.

A far rarer danger — and one that could not be avoided — is the blast of radiation from a nearby γ-ray burst. Perhaps the most frightening of these celestial explosions is the 'short-hard' γ-ray burst, caused by the violent merger of two black holes, two neutron stars or a combination. If one such blast were directed at Earth from within 200 parsecs away (less than 1% of the distance across the Milky Way), it would zap the globe with enough high-energy photons to wipe out 30% of the atmosphere's protective ozone layer for nearly a decade[5]. That sort of event — expected once every 300 million years or so — would double the amount of ultraviolet (UV) light reaching the ground and scorch phytoplankton, which make up the base of the ocean's food web.

Astronomers have no way of knowing whether such a rare event is imminent. Neutron stars are small and dark, so there is no catalogue of those within striking distance. "We wouldn't see it coming," says Thomas. In as-yet-unpublished work, he estimates that such an event could cause a 60% increase in UV damage to crops, with up to 60% reduction in crop yields.

From a distance of about 2,000 parsecs, 'long-soft' γ-ray bursts — which result from the collapse of massive stars — could also cause extinctions. But these events are rarer than short-hard bursts, and easier to spot in advance because they come from larger, brighter stars. The two-star system WR 104 is some 2,500 parsecs away from Earth, and is far enough along in its life cycle that it is expected to explode some time in the next few hundred thousand years — although the beam from the burst is unlikely to hit Earth.

It is possible that a γ-ray blast has hit the planet before. Melott, Thomas and their colleagues have suggested that the mass extinction at the end of the Ordovician period, 440 million years ago, could have been triggered by a γ-ray blast that wiped out some species through UV exposure and killed off others by creating a sunlight-blocking haze of nitrogen dioxide[6]. This would explain why some species went extinct before the globe cooled during that period, and it fits the extinction pattern, which shows that among marine organisms, the greatest toll was on plankton and other life in the upper part of the ocean.

Thomas says that none of these potential disasters is keeping him up at night. He does, however, "have some canned food in the basement" — a prudent backup in the event of any disaster.

DEATH BY WATER

Eight thousand years ago, sediments covering an underwater area the size of Scotland slipped from their moorings off the west coast of Norway and raced along the sea floor. The Storegga slide triggered a tsunami that ran at least 20 metres up the nearby Shetland Islands, and probably wiped out some coastal tribes as it clobbered shores around northern Europe. The scar it left on the ocean floor stretches nearly 300 kilometres. "It's absolutely enormous, and I'm not using the word 'enormous' lightly," says Peter Talling, a sedimentologist at the University of Southampton, UK, who is leading a project to assess the country's risk of similar slides.

The United Kingdom is not the only country concerned about giant submarine landslides. "There are definitely areas that have potential," says Uri ten Brink, a geophysicist at the US Geological Survey in Woods Hole, Massachusetts, who conducted a 2008 study of possible sources of tsunamis on the US east coast, where some nuclear power plants are within striking distance of such waves. "There are far larger piles of sediment around today than Storegga ever was," ten Brink says, including deposits along the coast of southern Alaska and off the Amazon, Niger and Nile river deltas. Smaller slides are more probable and can still have a huge local impact — and they often strike without warning. In 1998, a relatively small (magnitude-7) earthquake triggered an underwater slide that launched a 15-metre-high tsunami into Papua New Guinea, killing 2,200 people.

Researchers say that it is hard to quantify the threat of marine slides, particularly the giant ones. "There is so little information about events that happen so rarely," says ten Brink. "We just have to learn as much as we can." ■ SEE EDITORIAL P. 134

Nicola Jones *is a freelance writer near Vancouver, Canada.*

1. Mason, B. G., Pyle, D. M. & Oppenheimer, C. *Bull. Volcanol.* **66,** 735–748 (2004).
2. Isaia, R., Marianelli, P. & Sbrana, A. *Geophys. Res. Lett.* **36,** L21303 (2009).
3. Fisher, M. C. *et al. Nature* **484,** 186–194 (2012).
4. Haas, B. J. *et al. Nature* **461,** 393–398 (2009).
5. Melott, A. L. & Thomas, B. C. *Astrobiology* **11,** 343–361 (2011).
6. Melott, A. L. *et al. Int. J. Astrobiol.* **3,** 55–61 (2004).

COMMENT

ANTONIO SCORZA/AFP/GETTY

The shores of Rodrigo de Freitas Lake in Brazil are littered with fish killed by algal overgrowth in March 2000.

From patterns to predictions

Truly generic signals warning of tipping points are unlikely to exist, warn **Carl Boettiger** and **Alan Hastings**, so researchers should study transitions specific to real systems.

There has been much talk about tipping points over the past few years, and about the warning signals that may precede them. You could be forgiven for thinking that the forecasting of epidemics and stock-market crashes is just around the corner. But no one has yet managed to use the theory on early warning signals to predict a natural catastrophe.

The rewards of bridging the gap between the real world and mathematical conceptualizations of catastrophic shifts would be vast. Climate scientists might be able to foresee major shifts in the ocean currents with a rise in global temperatures; ecologists could potentially stave off pest outbreaks; and policies might be implemented to avert the collapse of fisheries[1]. (A report out this week from the World Economic Forum outlines other risks facing the world[2]). But for such applications to emerge, researchers should resist the lure of general rules. We must instead use all the available data to develop tools to study the specific properties of real systems.

The idea that critical transitions should be preceded by warning signals dates as far back as 1984 (ref. 3). In the past six years or so, such signals have been detected in various experiments. For instance, when yeast (*Saccharomyces cerevisiae*) is grown in chambers, the populations show some recovery after each drop in their nutrient supply as they adapt to the new conditions[4]. But this 'resettling' takes longer as the community approaches a crash. Also, the sizes of populations across chambers can become more variable in space and time[5] (see 'Predictors of doom').

Such warning signals are tantalizingly general. Researchers have seen the same dynamics — a slower return to equilibrium and increased variance — elsewhere[1], including in algal overgrowth in lakes, the end of ice ages (tracked by measurements ▶

SHATTIL & ROZINSKI/NPL

Forests in Colorado are being pushed towards collapse by pine-beetle infestations.

of the dust and atmospheric gas in ice cores) and the onset of epilepsy (signalled by a sudden change in the electrical activity of the brain).

However, no 'one-size-fits-all' property has been found that signals the imminent collapse of a complex system. Hypothetical systems can flip into different states without exhibiting either increased variance or a slower return to equilibrium[6,7]. And other models display exactly the opposite of what is expected. For instance, a model based on the life-history patterns of animals such as salamanders, which spend the juvenile and adult stages of their life cycles in different habitats, suggests that progressive habitat fragmentation can lead to a decrease in the variance of local population sizes[7].

Much effort is being dedicated to finding 'generic' warning signals that apply across diverse systems. But because the phenomena identified so far are not universally associated with tipping points, nor even sure indicators of a major shift, their predictive power is uncertain. We believe that in most cases, models designed to predict when critical transitions will happen, and in what circumstances, will need to be guided by — and perhaps even generated from — data on the specific system of interest.

IMPROVING PREDICTIONS

Encouragingly, even a small amount of system-specific information can dramatically improve researchers' ability to predict a collapse in simulations. For instance, by incorporating basic knowledge about the birth and death of North Atlantic cod into their models, ecologists were able to double the accuracy with which they predicted an approaching collapse of the population[8].

Also encouraging is that remote-sensing tools are flooding researchers with gigabytes of data[9]. Satellite images, for instance, provide detailed information about the health (and in some cases, species composition) of vast swathes of forests, including the Amazon rainforest and boreal forests. And electronic data loggers, such as continuous-monitoring floats, are providing detailed information on the physical and biological properties of the world's oceans.

The multidimensional data that these tools provide could lead to modelling approaches with the specificity and statistical power necessary for real-world applications. An as yet untested approach is to plug whatever data are available into machine learning algorithms to find out what, if any, warning signals manifest when certain parameters are adjusted.

Of course, to establish whether a candidate signal is outside the normal range, researchers need to know how the ecosystem, fault line, lake or whatever it is behaves under 'normal' circumstances. Yet controls or replicates are rarely available for natural systems.

One option is to compare the current properties of a system with historical or 'baseline' data. For such baselines to be effective, vast amounts of data must have been collected well before the system nears the tipping point.

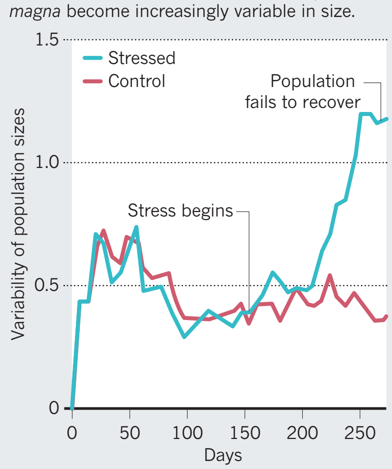
Such an approach needs exploring; we don't yet know whether it will even be possible to obtain baseline data for many complex systems. And for ecological systems, comparisons will often be confounded by undocumented changes in data-collection procedures, a switch in management practice or a shift in environmental conditions.

Another way around the lack of controls and replicates, and one that has received less attention, is to model (rather than measure) the expected behaviour of the stable system[10]. In this scenario, various data, including, say, real recordings of the density of invertebrates in a lake, could be fed into a model. In one simulation, a certain parameter, such as the number of predators, would be kept constant; in the other, it would be steadily increased to destabilize the food chain. By comparing various properties of the two systems, modellers could then establish the predictive power of candidate early warning signals.

Both theoretical[3] and experimental[4,5] approaches have shown that measurable signs of impending transitions do exist[1]. The challenge now is to resist the seductive but elusive concept of 'generic' warning signals, and focus instead on making predictions about real systems using readily available, or easily collectable, data. ■

PREDICTORS OF DOOM

Before crashing as a result of food-deprivation stress, populations of the water flea *Daphnia magna* become increasingly variable in size.



SOURCE: REF. 5

Carl Boettiger *is at the Center for Stock Assessment Research, Department of Applied Math and Statistics, University of California, Mail Stop SOE-2, Santa Cruz, California 95064, USA;* **Alan Hastings** *is at the Department of Environment Science & Policy, University of California, Davis, California 95616, USA.*
e-mail: cboettig@ucdavis.edu
amhastings@ucdavis.edu

1. Scheffer, M. *et al. Nature* **461,** 53–59 (2009).
2. World Economic Forum *Global Risks 2013* (in the press).
3. Wissel, C. *Oecologia* **65,** 101–107 (1984).
4. Dai, L., Vorselen, D., Korolev, K. S. & Gore, J. *Science* **336,** 1175–1177 (2012).
5. Drake, J. M. & Griffen, B. D. *Nature* **467,** 456–459 (2010).
6. Hastings, A. & Wysham, D. B. *Ecol. Lett.* **13,** 464–472 (2010).
7. Schreiber, S. & Rudolf, V. H. W. *Ecol. Lett.* **11,** 576–587 (2008).
8. Lade, S. J. & Gross, T. *PLoS Comput. Biol.* **8,** e1002360 (2012).
9. Dakos, V., Kéfi, S., Rietkerk, M., van Nes, E. H. & Scheffer, M. *Am. Nat.* **177,** E153–E166 (2011).
10. Boettiger, C. & Hastings, A. *J. R. Soc. Interface* **9,** 2527–2539 (2012).

Value all research products

A new funding policy by the US National Science Foundation represents a sea-change in how researchers are evaluated, says **Heather Piwowar**.

What a difference a word makes. For all new grant applications from 14 January, the US National Science Foundation (NSF) asks a principal investigator to list his or her research "products" rather than "publications" in the biographical sketch section. This means that, according to the NSF, a scientist's worth is not dependent solely on publications. Data sets, software and other non-traditional research products will count too.

There are more diverse research products now than ever before. Scientists are developing and releasing better tools to document their workflow, check each other's work and share information, from data repositories to post-publication discussion systems. As it gets easier to publish a wide variety of material online, it should also become easy to recognize the breadth of a scientist's intellectual contributions.

But one must evaluate whether each product has made an impact on its field — from a data set on beetle growth, for instance, to the solution to a colleague's research problem posted on a question-and-answer website. So scientists are developing and assessing alternative metrics, or 'altmetrics' — new ways to measure engagement with research output.

The NSF policy change comes at a time when around 1 in 40 scholars is active on Twitter[1], more than 2 million researchers use the online reference-sharing tool Mendeley (see go.nature.com/x63cwe), and more than 25,000 blog entries have been written about peer-reviewed research papers and indexed on the Research Blogging platform[2].

In the next five years, I believe that it will become routine to track — and to value — citations to an online lab notebook, contributions to a software library, bookmarks to data sets from content-sharing sites such as Pinterest and Delicious. In other words, to value a wider range of metrics that suggest a research product has made a difference. For example, my colleagues and I have estimated that the data sets added to the US National Center for Biotechnology Information's Gene Expression Omnibus in 2007 have contributed to more than 1,000 papers[3,4]. Such attributions continue to accumulate for several years after data sets are first made publicly available.

In the long run, the NSF policy change will do much more than just reward an investigator who has authored a popular statistics package, for instance. It will change the game, because it will alter how scientists assess research impact.

The new NSF policy states: "Acceptable products must be citable and accessible including but not limited to publications, data sets, software, patents, and copyrights." By contrast, previous policies allowed only "patents, copyrights and software systems" in addition to research publications in the biography section of a proposal, and considered their inclusion to be a substitute for the main task of listing research papers.

Still, the status quo is largely unchanged. Some types of NSF grant-renewal applications continue to request papers alone. Indeed, several funders — including the US National Institutes of Health, the Howard Hughes Medical Institute and the UK Medical Research Council — still explicitly ask for a list of research papers rather than products.

> ***"Altmetrics give a fuller picture of how research products have influenced conversation, thought and behaviour."***

Even when applicants are allowed to include alternative products in grant applications, how will reviewers know if they should be impressed? They might have a little bit of time to watch a short video on YouTube demonstrating a wet-lab technique, or to read a Google Plus post describing a computational algorithm. But what if the technique takes more time to review, or is in an area that is outside the reviewer's expertise? Existing evaluation mechanisms often fail for alternative products — a YouTube video, for example, has no journal title to use as a proxy for anticipated impact. But it will definitely receive a number of downloads, some 'likes' on Facebook, a few Pinterest bookmarks and discussion in blogs.

TRACKING TRENDS

Many altmetrics have already been gathered for a range of research products. For example, the data repositories Dryad and figshare track download statistics (figshare is supported by Digital Science, which is owned by the same parent company as *Nature*). Some repositories, such as the Inter-university Consortium for Political and Social Research, provide anonymous demographic breakdowns of usage.

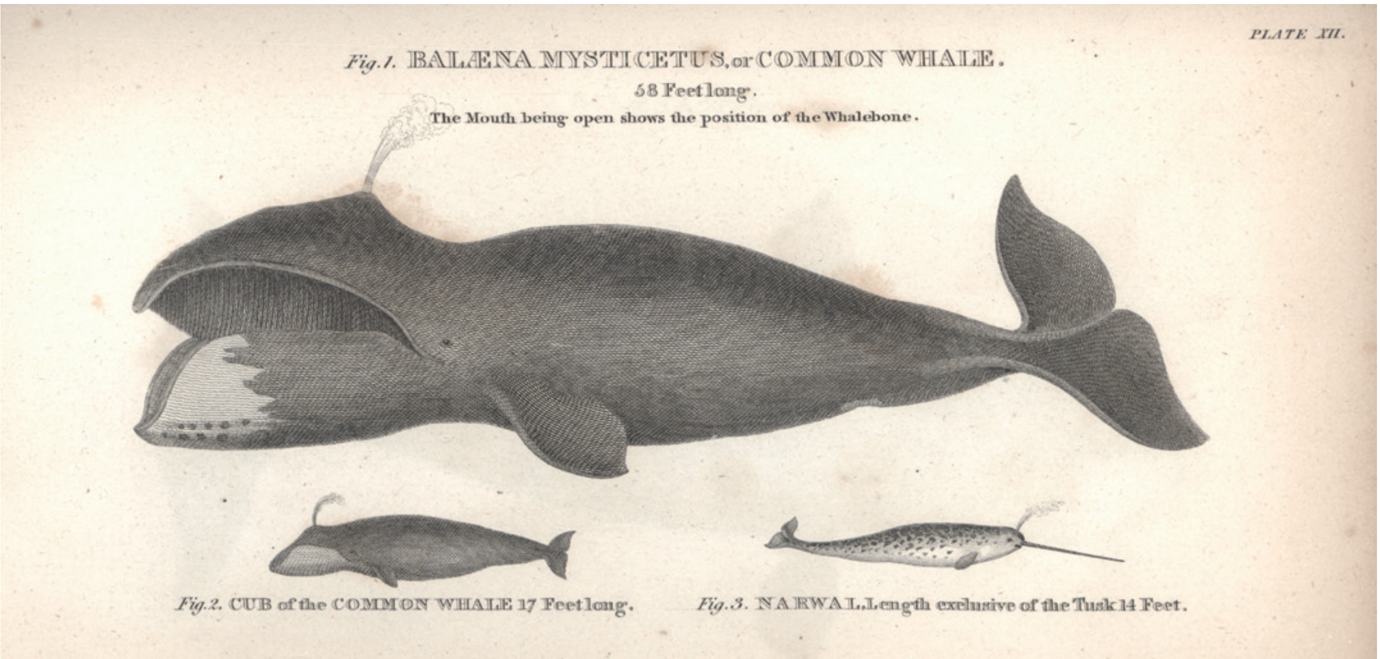
Specific tools have been built to aggregate altmetrics across a wide variety of content. Altmetric.com (also supported by Digital Science) reveals the impact of anything with a digital object identifier (DOI) or other standard identifier. It can find mentions of a data set in blog posts, tweets and mainstream media (see go.nature.com/yche8g). The non-profit organization ImpactStory (http://impactstory.org), of which I am a co-founder, tracks the impact of articles, data sets, software, blog posts, posters and lab websites by monitoring citations, blogs, tweets, download statistics and attributions in research articles, such as mentions within methods and acknowledgements[5]. For example, a data set on an outbreak of *Escherichia coli* has received 43 'stars' in the GitHub software repository, 18 tweets and two mentions in peer-reviewed articles (see go.nature.com/dnhdgh).

Such altmetrics give a fuller picture of how research products have influenced conversation, thought and behaviour. Tracking them is likely to motivate more people to release alternative products — scientists say that the most important condition for sharing their data is ensuring that they receive proper credit for it[6].

The shift to valuing broad research impact will be more rapid and smooth if more funders and institutions explicitly welcome evidence of impact. Scientists can speed the shift by publishing diverse research products in their natural form, rather than shoehorning everything into an article format, and by tracking and reporting their products' impact. When we, as scientists, build and use tools and infrastructure that support open dissemination of actionable, accessible and auditable metrics, we will be on our way to a more useful and nimble scholarly communication system. ■

Heather Piwowar *is a postdoctoral research associate in informatics at the National Evolutionary Synthesis Center, Duke University, Durham, North Carolina, USA. She is a co-founder of ImpactStory.*
e-mail: heather@impactstory.org

1. Priem, J., Costello, K. & Dzuba, T. *Figshare* http://dx.doi.org/10.6084/m9.figshare.104629 (2012).
2. Fausto, S. *et al. PLoS ONE* **7,** e50109 (2012).
3. Piwowar, H. A., Vision, T. J. & Whitlock, M. C. *Nature* **473,** 285 (2011).
4. Piwowar, H. A., Vision, T. J. & Whitlock, M. C. *Dryad Digital Repository* http://dx.doi.org/10.5061/dryad.j1fd7 (2011).
5. Oettl, A. *Nature* **489,** 496–497 (2012).
6. Tenopir, C. *et al. PLoS ONE* **6,** e21101 (2011).



NOAA

This 1820 plate is one of a number by William Scoresby, a whaler's son who was the first person to draw and describe whales accurately.

CETOLOGY

How science inspired *Moby-Dick*

Philip Hoare tracks the scientific influences and insights that breach throughout Herman Melville's epic novel.

More than a century and a half after it was published, Herman Melville's *Moby-Dick* remains a key cultural bridge between human history and natural history — expressed in the vast and ominous shape of the whale. This epic novel is a laboratory of literature, created in an age before art and science became strictly demarcated.

Melville wrote his book — which drew on his own youthful experiences on a whaling ship — as a tribute to the first period of modern whaling in the eighteenth to mid-nineteenth centuries, which he claimed to be worth US$7 million a year to the fledgling United States. At the same time, science was undergoing a sea change as the gentleman scientists and polymaths of the century's start gave way to more specialized and professionalized successors.

Melville's attitude to, and use of, science in *Moby-Dick* was in line with the eclectic ethos of that period. Drawing on the work of luminaries such as William Scoresby, Thomas Beale, Georges Cuvier and Louis Agassiz, Melville used contemporary knowledge of natural history — or the lack of it — to his own ends.

Seventeen of the book's 135 chapters focus on whale anatomy or behaviour. Titles include 'The Sperm Whale's Head — Contrasted View' and 'The Right Whale's Head — Contrasted View'; such sections lay out the whales' physical structure with a wry mixture of known facts and arch analogy. (In a witty 2011 essay, marine biologist Harold Morowitz speculates on Melville as a "cetacean gastroenterologist or proctologist".) Melville's must also be the first, and perhaps last, work of literature to feature a chapter on zooplankton.

Moby-Dick
HERMAN MELVILLE
Harper & Brothers: 1851.

In the famous Chapter 32, 'Cetology', Melville attempts to categorize species of whale as he would catalogue his library, in 'folios'. It was a playful gesture that reflected the fluid classification of cetacean species at the time. In *The Natural History of the Sperm Whale* (1839), Beale notes that the French natural historian Bernard Germaine de Lacépède claimed that there were eight species of this whale; there are in fact only three: *Physeter macrocephalus; Kogia sima*, the dwarf sperm whale; and the pygmy, *K. breviceps*. Accordingly, Melville pronounces earlier attempts to describe whales "all wrong", and deploys two contemporary authorities to bolster his claim: Scoresby and Beale.

➲ **NATURE.COM**
For more on the scientific history of whaling, see:
go.nature.com/7ykg2n

The Natural History of the Sperm Whale was the first attempt to write scientifically about this deep-diving, open-ocean whale. The result of Beale's experiences as a surgeon on a British whaling ship, the book was full of observations on the animal's anatomy and behaviour. Cuvier had claimed that the sperm whale struck fear into "all the inhabitants of the deep", but Beale knew this whale to be "a most timid and inoffensive animal".

Equally, Scoresby's groundbreaking *An Account of the Arctic Regions* (1820) gave Melville insight into the other cetacean whose numbers were decimated by whaling: the bowhead (*Balaena mysticetus*), then known as the common whale. Scoresby, the son of a whaler, was a typical polymath of the time: hunter, scientist, clergyman and mesmerist. In his early career he had received encouragement from Joseph Banks, and his work set the benchmark for Arctic studies.

Melville was particularly fascinated by Scoresby's observations of an ancient Inuit harpoon embedded in a bowhead's blubber. "Who had darted that stone lance?" Melville's narrator Ishmael wonders, imagining (with slight exaggeration) that it had been thrown "long before America was

discovered". Science indicates that Melville may not have been far wrong. In 1999, tests on bowheads indicated that these animals can live for at least 200 years.

Of course, the greatest scientific figure of the age hovers over Melville. Darwin published *On the Origin of Species* in 1859, eight years after *Moby-Dick* came out. Melville's sole mention of Darwin is a quote — from *Darwin's Voyage of a Naturalist* (sic) — in the extracts at the start of *Moby-Dick*. He had read Darwin's *Voyage of the Beagle* (1839) in preparation for his own 1854 work, *The Encantadas or Enchanted Isles* — as the Galapagos were then known. Melville visited the islands in 1841, six years after Darwin's fateful landing. Darwin's recorded observation of marine iguanas as "imps of darkness" seemed to set the tone for Melville's metaphoric view of the Galapagos, which he saw as "five-and-twenty heaps of cinders … In no world but a fallen one could such lands exist".

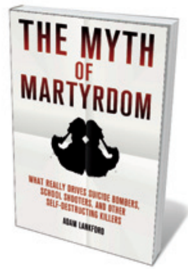
Such dark analogies are in line with a man who declared all human science to be "but a passing fable" — and yet created a fable of his own. In *Moby-Dick*, Ishmael is a perpetually sceptical and questioning figure, a man attuned to science — a stark contrast to the vengeful Ahab and his pursuit of the whale that "dismasted" him. As the critic Eric Wilson, in his essay 'Melville, Darwin, and the Great Chain of Being', notes, a "primary subtext of Melville's novel is the passing of pre-Darwinian, anthropocentric thought, espoused by Ahab, and the inauguration of a version of Darwin's more ecological evolution, proffered by Ishmael".

"Melville's masterpiece resonates powerfully with today's scientific concerns."

Melville lived through that process. US Transcendentalist Ralph Waldo Emerson's essay *Nature* (1836), with its declaration of moral law at the heart of the cosmos, was the new philosophy of Melville's youth. But as biographer Andrew Delbanco points out, Melville read *A Hazard of New Fortunes* (1890), William Dean Howells's Darwinian-inflected view of society. *Moby-Dick* itself has been seen as a parody of the Transcendentalists' 'back-to-nature' excesses. But Melville does more than lambast philosophy or use science as interior decoration. He achieved a marvellous synthesis of his own poetic and philosophical impulse with the increasingly science-aware ethos of his age. And he did so with a sense of black humour that transcended Transcendentalism to prove that nature — and its science — was much stranger and more wonderful than they had imagined.

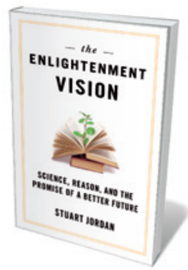
Moby-Dick failed to make any impact in Melville's lifetime, and he died forgotten in 1891. But his spirit of enquiry and experiment stood him in good stead as far as ▶

Books in brief

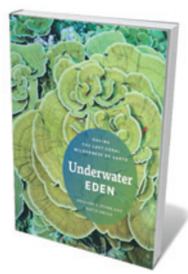


The Myth of Martyrdom: What Really Drives Suicide Bombers, Rampage Shooters, and Other Self-Destructive Killers
Adam Lankford PALGRAVE MACMILLAN *272 pp. £16.99 (2013)*
Are suicide bombers psychologically normal? Many psychologists, including experts 'diagnosing' the hijackers responsible for the 11 September 2001 terrorist attacks, view them as just that, albeit exercised by a powerful sense of justice. Adam Lankford begs to differ. Self-destructive killers, he says, are already primed for suicide — so depressed, addicted or brutalized that it is relatively easy to tip them over the edge. A criminal-justice specialist, Lankford presents compelling, well-synthesized evidence for his case.

The White Planet: The Evolution and Future of Our Frozen World
Jean Jouzel, Claude Lorius and Dominique Raynaud. Translated by Teresa Lavender Fagan PRINCETON UNIV. PRESS *316 pp. $29.95 (2013)*
Ice in all its chill Earthly manifestations has drawn thousands of research scientists into the white deserts of the world. Now, three pioneers of ice-core science — Jean Jouzel, Claude Lorius and Dominique Raynaud — reveal key facets of the cryosphere in a new translation of their sweeping overview. Moving from exploration and early science, they delve into the ice 'archives' and findings on climate ancient and current, the rise of pollution and more. A nuanced and thorough look at climate change and its implications.



The Enlightenment Vision: Science, Reason, and the Promise of a Better Future
Stuart Jordan PROMETHEUS *295 pp. $26 (2013)*
Physicist Stuart Jordan scrutinizes the afterglow of that scientific big bang, the Enlightenment of the seventeenth and eighteenth centuries. Aspects of today's culture — medicine, scientific outlook, democracy and technological advances — carry traces of the original vision. But Jordan shows too how mixed a legacy we face, from ignorance about science, a bulging population and "juggernaut technology" to degraded ecosystems. Particularly by upholding ethics, he argues, we can collectively turn the tide.



Underwater Eden: Saving the Last Coral Wilderness on Earth
Gregory S. Stone and David Obura UNIV. CHICAGO PRESS *184 pp. $40 (2012)*
Ocean warming and acidification are bad news for corals, and more than one-quarter of fish species. So when Gregory Stone dived around the remote Pacific Phoenix Islands in 2002, he was stunned to see a 'lost world' of untouched coral beds. Here Stone, chief ocean scientist of Conservation International, coral researcher David Obura and contributors lay out what happened next: the hard-won creation of the largest World Heritage Site ever sanctioned by the United Nations Educational, Scientific and Cultural Organization.

This Explains Everything: Deep, Beautiful, and Elegant Theories of How the World Works
John Brockman HARPER PERENNIAL *432 pp. $15.99 (2013)*
Agent to the stars of science, John Brockman presents mind-bites from his stable of research heavyweights asked to name their "favourite deep, elegant, or beautiful explanation". Try theoretical physicist Freeman Dyson speculating on the putative coexistence of quantum and classical world views, or mathematician Samuel Arbesman admiring the reaction–diffusion model that dictates a leopard's spots.

TIME LIFE PICTURES/GETTY

Moby-Dick **author Herman Melville.**

literary immortality is concerned. His allusive style chimed with a new century of discovery, and twentieth-century experimentalists of literature such as D. H. Lawrence and Virginia Woolf reappraised him as a modernist who lived before modernism was invented.

Melville's masterpiece also resonates powerfully with today's scientific concerns. *Moby-Dick* contrasts the glory of the whale with the threats posed by humanity. Melville even seems to anticipate the effects of a changing environment. In the moving chapter 'Does The Whale's Magnitude Diminish? — Will He Perish?', Melville wonders about a flooded future, but sees the whale as triumphant, spouting "his frothed defiance to the skies". Yet by the time his book finally came into its own, Melville's vision had turned into a nightmare for the whale.

In 1961 alone, more whales died — nearly 75,000 — than in the entire span of Yankee whaling. With faster ships and grenade harpoons, new species had come within the hunters' remit: the blue and fin whales of the South Atlantic and Southern Ocean. And, like Scoresby, the "hip-booted cetologists" (as D. Graham Burnett describes them in his *The Sounding of the Whale: Science and Cetaceans in the Twentieth Century*, University of Chicago Press, 2012), entered a complicit arrangement with the modern whaling industry to inform their conclusions on whale anatomy, breeding and migration. It is telling, perhaps, that no one has written a follow-up to *Moby-Dick* to celebrate that particular adventure. ■

Philip Hoare *is the author of* Leviathan, or, The Whale. *His new book,* The Sea Inside, *is due out in June. He is co-curator of the* Moby-Dick *Big Read (www.mobydickbigread.com), an online project hosted by Plymouth University, UK.*
e-mail: phoare@themovement.demon.co.uk

INNOVATION

Motley inventors

John Browning welcomes a collective portrait of creators that puts life stories first.

An enthusiast's book about enthusiasts, *The Tinkerers* surveys a motley collection of US innovators whose creations are changing the world — or so their makers hope. The result is a kaleidoscopic view of the myriad forms innovation can take.

Alec Foege's book is a useful contribution to understanding our era, repeatedly transformed by innovation that has generated pages and gigabytes of analysis in the abstract, but surprisingly little on the inventors themselves. Foege redresses the balance, covering a crew that ranges from Thomas Edison to Silicon Valley whizz-kids. He tries to let his inventors describe in their own words what they thought they were doing, and why.

Dean Kamen, inventor of the Segway and various medical devices, falls closest to the mould set by Edison. Based in New Hampshire, he has used the cash flow from licensing his automatic syringe, dialysis machine and other medical innovations to invent solutions to whatever takes his fancy — such as a wheelchair that can navigate steps. Saul Griffith puts a more modern spin on the model. His "do tank", Squid Labs in Alameda, California, spins off companies rather than licences — including Howtoons, a website featuring educational cartoons, and Makani Power, which uses airborne wind turbines to harvest the energy of high-altitude winds.

Meanwhile, Nathan Myhrvold, Microsoft's former head of technology, takes new models of innovation to a logical extreme. Instead of creating his own inventions, his company, Intellectual Ventures, buys, finances and creates a large patent portfolio, which it then licenses to others. He argues that this inspires a broad swathe of creativity. Some, in turn, argue that it encourages a profusion of patents and could merely mire innovation in legal uncertainties and lawsuits.

Foege doesn't provide any solutions, or even take a strong stance in that debate. For better or worse, *The Tinkerers* feels like a prototype — a bit rough and ready, created as much to point towards interesting questions as to provide answers. To be truly representative, the choice of profiles should at least have included some innovators in biology or medical science, and some working in corporate laboratories. But the hotchpotch does capture a range of approaches and motivations. Kamen seems to like solving problems. Griffith wants to create a greener, smarter world. And although he contributed to the underpinnings of Microsoft Windows, Myrhvold now sees his role in promoting innovation as largely financial and legal.

The Tinkerers: The Amateurs, DIYers, and Inventors Who Make America Great
ALEC FOEGE
Basic Books: 2013. 224 pp. $26.99

Foege's approach contrasts with the books on innovation that take the 'grand theory of creativity' route, knocking the sharp edges off individual histories to fit them in. Steven Johnson's *Where Good Ideas Come From* (Riverhead, 2010) fits inspiration into seven categories, including hunches and happy accidents. Clayton Christensen's *The Innovator's Dilemma* (Harvard Business School Press, 1997) divvies up the world of innovation according to its impact on business strategy. Fascinating though such books often are, this abstract point of view was more valuable 50 years ago than today.

Then, innovation was managed. Much of it happened in government-funded set-ups such as Bell Labs and DARPA (the US Defense Advanced Research Projects Agency), or at big corporate-funded labs like IBM and Dow Chemical. President Dwight D. Eisenhower, in his 1961 farewell address, worried that "a government contract becomes virtually a substitute for intellectual curiosity". And so powerful were the managers of that funding that he followed his warning about the military–industrial complex with another about the "scientific-technological elite".

Nearly 30 years ago, Steven Levy's book *Hackers* (Anchor/Doubleday, 1984) inspired a generation with the idea that ideas born of subversion are often more powerful than those blessed by management. Today, innovation is popping up all over the place: garages; 'hack-spaces'; bedrooms. Nobody is managing the process. So everybody's story matters.

Few of those stories have yet been told, making those that have the exceptions. One example, Walter Isaacson's biography of Steve Jobs, sparked a global debate about how to manage innovation grounded in life rather than theory. But the exceptions should become the rule. Innovators' histories need to be written. For all its flaws, Foege's book is a step in the right direction. ■

John Browning *is a freelance journalist and consultant. He was formerly an editor at* The Economist *and* Wired.
e-mail: jb@poplar.com

TIME LIFE PICTURES/GETTY

Moby-Dick **author Herman Melville.**

literary immortality is concerned. His allusive style chimed with a new century of discovery, and twentieth-century experimentalists of literature such as D. H. Lawrence and Virginia Woolf reappraised him as a modernist who lived before modernism was invented.

Melville's masterpiece also resonates powerfully with today's scientific concerns. *Moby-Dick* contrasts the glory of the whale with the threats posed by humanity. Melville even seems to anticipate the effects of a changing environment. In the moving chapter 'Does The Whale's Magnitude Diminish? — Will He Perish?', Melville wonders about a flooded future, but sees the whale as triumphant, spouting "his frothed defiance to the skies". Yet by the time his book finally came into its own, Melville's vision had turned into a nightmare for the whale.

In 1961 alone, more whales died — nearly 75,000 — than in the entire span of Yankee whaling. With faster ships and grenade harpoons, new species had come within the hunters' remit: the blue and fin whales of the South Atlantic and Southern Ocean. And, like Scoresby, the "hip-booted cetologists" (as D. Graham Burnett describes them in his *The Sounding of the Whale: Science and Cetaceans in the Twentieth Century*, University of Chicago Press, 2012), entered a complicit arrangement with the modern whaling industry to inform their conclusions on whale anatomy, breeding and migration. It is telling, perhaps, that no one has written a follow-up to *Moby-Dick* to celebrate that particular adventure. ■

Philip Hoare *is the author of* Leviathan, or, The Whale. *His new book,* The Sea Inside, *is due out in June. He is co-curator of the* Moby-Dick *Big Read (www.mobydickbigread.com), an online project hosted by Plymouth University, UK.*
e-mail: phoare@themovement.demon.co.uk

INNOVATION

Motley inventors

John Browning welcomes a collective portrait of creators that puts life stories first.

An enthusiast's book about enthusiasts, *The Tinkerers* surveys a motley collection of US innovators whose creations are changing the world — or so their makers hope. The result is a kaleidoscopic view of the myriad forms innovation can take.

Alec Foege's book is a useful contribution to understanding our era, repeatedly transformed by innovation that has generated pages and gigabytes of analysis in the abstract, but surprisingly little on the inventors themselves. Foege redresses the balance, covering a crew that ranges from Thomas Edison to Silicon Valley whizz-kids. He tries to let his inventors describe in their own words what they thought they were doing, and why.

Dean Kamen, inventor of the Segway and various medical devices, falls closest to the mould set by Edison. Based in New Hampshire, he has used the cash flow from licensing his automatic syringe, dialysis machine and other medical innovations to invent solutions to whatever takes his fancy — such as a wheelchair that can navigate steps. Saul Griffith puts a more modern spin on the model. His "do tank", Squid Labs in Alameda, California, spins off companies rather than licences — including Howtoons, a website featuring educational cartoons, and Makani Power, which uses airborne wind turbines to harvest the energy of high-altitude winds.

Meanwhile, Nathan Myhrvold, Microsoft's former head of technology, takes new models of innovation to a logical extreme. Instead of creating his own inventions, his company, Intellectual Ventures, buys, finances and creates a large patent portfolio, which it then licenses to others. He argues that this inspires a broad swathe of creativity. Some, in turn, argue that it encourages a profusion of patents and could merely mire innovation in legal uncertainties and lawsuits.

Foege doesn't provide any solutions, or even take a strong stance in that debate. For better or worse, *The Tinkerers* feels like a prototype — a bit rough and ready, created as much to point towards interesting questions as to provide answers. To be truly representative, the choice of profiles should at least have included some innovators in biology or medical science, and some working in corporate laboratories. But the hotchpotch does capture a range of approaches and motivations. Kamen seems to like solving problems. Griffith wants to create a greener, smarter world. And although he contributed to the underpinnings of Microsoft Windows, Myrhvold now sees his role in promoting innovation as largely financial and legal.

The Tinkerers: The Amateurs, DIYers, and Inventors Who Make America Great
ALEC FOEGE
Basic Books: 2013. 224 pp. $26.99

Foege's approach contrasts with the books on innovation that take the 'grand theory of creativity' route, knocking the sharp edges off individual histories to fit them in. Steven Johnson's *Where Good Ideas Come From* (Riverhead, 2010) fits inspiration into seven categories, including hunches and happy accidents. Clayton Christensen's *The Innovator's Dilemma* (Harvard Business School Press, 1997) divvies up the world of innovation according to its impact on business strategy. Fascinating though such books often are, this abstract point of view was more valuable 50 years ago than today.

Then, innovation was managed. Much of it happened in government-funded setups such as Bell Labs and DARPA (the US Defense Advanced Research Projects Agency), or at big corporate-funded labs like IBM and Dow Chemical. President Dwight D. Eisenhower, in his 1961 farewell address, worried that "a government contract becomes virtually a substitute for intellectual curiosity". And so powerful were the managers of that funding that he followed his warning about the military–industrial complex with another about the "scientific-technological elite".

Nearly 30 years ago, Steven Levy's book *Hackers* (Anchor/Doubleday, 1984) inspired a generation with the idea that ideas born of subversion are often more powerful than those blessed by management. Today, innovation is popping up all over the place: garages; 'hack-spaces'; bedrooms. Nobody is managing the process. So everybody's story matters.

Few of those stories have yet been told, making those that have the exceptions. One example, Walter Isaacson's biography of Steve Jobs, sparked a global debate about how to manage innovation grounded in life rather than theory. But the exceptions should become the rule. Innovators' histories need to be written. For all its flaws, Foege's book is a step in the right direction. ■

John Browning *is a freelance journalist and consultant. He was formerly an editor at* The Economist *and* Wired.
e-mail: jb@poplar.com

QUANTUM PHYSICS

Time crystals

Physicists have come up with the mind-boggling concept of a time crystal. This intriguing proposal, which is based on the notion of broken time-translation symmetry, might open up a whole new field of research.

PIERS COLEMAN

Philosophers, physicists and writers have long pondered the perpetual periodicities in time that make up our lives — from a heartbeat to the motions of the planets. Isaac Newton and Leonardo da Vinci both considered perpetual motion, concluding that it is the stuff of alchemy. Curiously, we now know that perpetual motion is possible, but perpetual oscillation, that's another thing. There is a new debate afoot involving the connection between time and 'broken symmetry'. A snowflake epitomizes this concept, because as it crystallizes, it lowers its energy by breaking the homogeneity and isotropy of water. Such 'spontaneously broken symmetry' permeates the world we live in: for example, the orientation of the liquid crystals in a television screen and the tiny range of forces between subatomic particles both rely on broken symmetry. These cases of broken symmetry leave the homogeneity of time unchallenged, but the world we live in is also full of clocks — of periodicities and oscillations that break time-translation symmetry.

A series of three papers[1–3] published in *Physical Review Letters* proposes and explores the evocative new concept of a time crystal. The authors of these papers ask: can temporal periodicities spontaneously develop in a fashion analogous to the spatial order of crystals? Naturally oscillating physical systems abound: for example, the sound of an organ pipe and the workings of clocks in all their variety. But these systems require an input of energy to drive the oscillation. The idea behind a crystal in the time domain is that it requires no energy input. Like its spatial cousin, it will spontaneously form without a driving force.

It turns out that the concept of a time crystal requires grappling with various subtle issues in quantum mechanics. In quantum mechanics, states of definite energy are called stationary states because the rate of change of any observable quantity associated with the state is zero. Paradoxically, this does not preclude states of perpetual motion: a famous example is superconductivity, whereby certain metals, when cooled, support perpetual electric currents. It is these currents that drive the magnetic fields in a magnetic resonance imaging scanner. If a tiny magnetic field is passed through a metal ring, which is then cooled into the superconducting state, the ring spontaneously develops a persistent current. But whereas the spontaneously developed motion is perpetual, the current is homogeneous and constant, so time-translation symmetry is still unbroken.

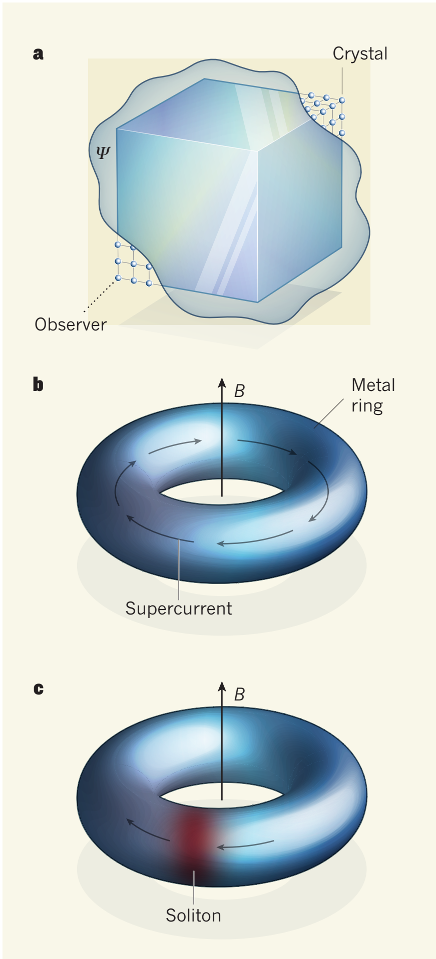


Figure 1 | Periodicities in space and time. a, The wavefunction, Ψ, of a crystal is spatially homogeneous. But when observed at one corner, it is revealed to have infinite-range spatial correlations associated with broken symmetry, exhibiting a type of order called off-diagonal long-range order[4]. **b**, A tiny magnetic field, B, introduced into a metal ring induces a persistent 'supercurrent' in the ring, but time-translation symmetry is unbroken. **c**, Inducing a weak attraction between the particles that carry the supercurrent causes them to bunch together, leading to the development of a 'soliton' in the particle density. Wilczek proposes[1] that, if this soliton is carried around with the superflow in the presence of a small magnetic field, time-translation symmetry is broken and the rotating soliton will form a periodic structure in time — a time crystal.

In physics, the idea of broken symmetry is associated with the concept of an 'order parameter': in a crystal, this is the periodic modulation in its density. Broken symmetry manifests itself through the development of infinite-range correlations of the order parameter. Thus, in a crystal, the density modulates coherently over infinite distances. In quantum physics, broken symmetry is not immediately evident from a system's wavefunction, because the vertices of the crystal are equally likely to occupy any point in space. For example, you cannot see the broken symmetry in a crystal until you measure the position of one of its atomic constituents. The corresponding crystal wavefunction is completely homogeneous in space, because it is a superposition of broken-symmetry states; once you observe a small piece of the crystal, the modulated density becomes visible. The development of these entangled, infinite-range spatial correlations in an otherwise spatially homogeneous wavefunction is called off-diagonal long-range order[4].

In the first of the three papers, Wilczek introduces[1] the concept of a time crystal, arguing that such a state would exhibit off-diagonal long-range order in the time domain. Wilczek proposes a thought experiment involving the supercurrent spontaneously induced in a superconducting ring threaded by a tiny magnetic field. Suppose, he argues, one were to induce a weak attraction between the particles in the superconductor, causing them to bunch together. Such a state is called a soliton. Provided that the soliton forms without eliminating the superconductivity, then in a small magnetic field a supercurrent will spontaneously emerge, carrying the soliton periodically around the ring to form a simple example of a time crystal (Fig. 1).

Although crystals owe their microscopic physics to quantum mechanics, their macroscopic behaviour is governed by classical mechanics: time crystals are surely no

different. In the second paper, Shapere and Wilczek show[2] that, to understand time crystals, one needs an advanced version of classical mechanics based on Lagrange's 'least action' principle, in which the Newtonian motion of particles is understood as the path that minimizes a combination of kinetic and potential energy called the action. They find that time crystals are classically feasible, but discover that, curiously, once a time crystal develops, the concept of energy becomes ambiguous and the action principle takes its place.

But if time crystals exist, how do we detect them? In the last of the three papers, Li and collaborators[3] suggest a method to engineer a time crystal using an ion trap. When ions are trapped in a circular ring, their repulsive interactions cause them to arrange themselves into a spatial crystal. Following Wilczek's proposal[1], Li *et al.* suggest that, when cooled in a small magnetic field, the crystal will spontaneously rotate. This idea is closely related to the idea of 'supersolids', itself a subject of great recent interest and controversy[5].

Perhaps the most important aspect of the time-crystal notion is that it envisages broken time-translation symmetry as an equilibrium phenomenon, rather than a non-equilibrium response to a driving force. This is an exciting yet controversial[6] development, which if right will open up a whole new field of research. With the discovery of superconductors, the debate on perpetual motion was brought to an end. With time crystals, a new debate has just begun. ■

Piers Coleman *is in the Center for Materials Theory, Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854–8019, USA.*
e-mail: coleman@physics.rutgers.edu

1. Wilczek, F. *Phys. Rev. Lett.* **109,** 160401 (2012).
2. Shapere, A. & Wilczek, F. *Phys. Rev. Lett.* **109,** 160402 (2012).
3. Li, T. *et al. Phys. Rev. Lett.* **109,** 163001 (2012).
4. Yang, C. N. *Rev. Mod. Phys.* **34,** 694–704 (1962).
5. Voss, D. *Physics* **5,** 111 (2012).
6. Bruno, P. Preprint at http://arxiv.org/abs/1210.4128 (2012).

ANIMAL BEHAVIOUR

Older but less wise

Most fish living in marine reserves are older, bigger and more fecund than those outside their borders, but they are also slower to flee a threat. The potential for 'spillover' of such fish into fisheries may boost support for reserves.

PETER F. SALE

Marine reserves are designed to protect the aquatic life within their margins. It seems probable that, by providing a place for fish to grow and reproduce, they will also benefit adjacent fisheries. Writing in *Ecology Letters*, Januchowski-Hartley *et al.*[1] add a new twist to this concept, by demonstrating that fish that wander over a reserve boundary into waters in which fishing occurs are more easily caught than those that have lived continuously with the threat of fishing.

Fish living in a reserve in which fishing is prohibited are expected to live longer, grow larger and have more offspring than those not so protected[2]. Some of these offspring will disperse naturally beyond the boundaries as eggs or larvae, in a process called recruitment subsidy; the remaining offspring generate greater fish numbers and biomass inside the reserve. This should, in turn, lead to 'spillover' — the continuous net movement of juvenile and adult individuals beyond reserve boundaries. The effects of recruitment subsidy are thought to be large and widespread because the larvae of most fish species are pelagic (they live within the water column, rather than near the bottom) and are therefore readily dispersed. By contrast, the spillover effect is expected to be restricted to sites close to the reserve boundary.

However, this eminently logical theory depends on assumptions that are seldom stated explicitly[2]. For this model to hold, the reserve must be large enough (or the fish sedentary enough) for individuals to experience reduced fishing pressure — a fish that wanders widely, swimming in and out of a reserve, will get little benefit. The reserve must also be well managed, so that fishing pressure within its boundaries really is lower than that outside. And the fish must produce offspring that are dispersive early in life. If these requirements are met, marine reserves should enhance fishery yields through both recruitment subsidy and spillover, and this expectation has become a commonly used argument in attempts to convince fishing communities to give up part of their fishing area for the establishment of a marine reserve.

The enthusiasm for marine reserves in the conservation community has led to strong advocacy for their use as a fishery-management tool — an advocacy effort that has jumped well ahead of the evidence that real augmentation of fishing actually results. Although the extent of this overreach has been noted several times[2,3], data showing that some reserves seem not to enhance fisheries[4] have been glossed over, and less-positive reports, such as one showing that the increase in marine protected areas worldwide is having no effect on the global loss of biodiversity[5], can have somewhat difficult paths to publication. Nevertheless, despite the hype, there is now abundant evidence that, for reasonably sedentary species in many environments, marine reserves will contain the more abundant and older fish expected by theory[2,4]. Concurrently, there is growing evidence of

REINHARD DIRSCHERL/FLPA

Figure 1 | Beyond the boundaries. Januchowski-Hartley *et al.*[1] compared the behaviour of fish species that are targets for fishing, including surgeonfish (such as *Naso vlamingii*, shown) and parrotfish, with non-targeted butterflyfish. They show that the flight initiation distance — the distance between a fish and an approaching snorkeller at which the fish flees — is smaller for targeted species in marine reserves than for those in fisheries. Moreover, this distance increases with increasing distance from the reserve boundary. No difference in flight behaviour was observed in the non-targeted species.

different. In the second paper, Shapere and Wilczek show[2] that, to understand time crystals, one needs an advanced version of classical mechanics based on Lagrange's 'least action' principle, in which the Newtonian motion of particles is understood as the path that minimizes a combination of kinetic and potential energy called the action. They find that time crystals are classically feasible, but discover that, curiously, once a time crystal develops, the concept of energy becomes ambiguous and the action principle takes its place.

But if time crystals exist, how do we detect them? In the last of the three papers, Li and collaborators[3] suggest a method to engineer a time crystal using an ion trap. When ions are trapped in a circular ring, their repulsive interactions cause them to arrange themselves into a spatial crystal. Following Wilczek's proposal[1], Li *et al.* suggest that, when cooled in a small magnetic field, the crystal will spontaneously rotate. This idea is closely related to the idea of 'supersolids', itself a subject of great recent interest and controversy[5].

Perhaps the most important aspect of the time-crystal notion is that it envisages broken time-translation symmetry as an equilibrium phenomenon, rather than a non-equilibrium response to a driving force. This is an exciting yet controversial[6] development, which if right will open up a whole new field of research. With the discovery of superconductors, the debate on perpetual motion was brought to an end. With time crystals, a new debate has just begun. ■

Piers Coleman *is in the Center for Materials Theory, Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854–8019, USA.*
e-mail: coleman@physics.rutgers.edu

1. Wilczek, F. *Phys. Rev. Lett.* **109,** 160401 (2012).
2. Shapere, A. & Wilczek, F. *Phys. Rev. Lett.* **109,** 160402 (2012).
3. Li, T. *et al. Phys. Rev. Lett.* **109,** 163001 (2012).
4. Yang, C. N. *Rev. Mod. Phys.* **34,** 694–704 (1962).
5. Voss, D. *Physics* **5,** 111 (2012).
6. Bruno, P. Preprint at http://arxiv.org/abs/1210.4128 (2012).

ANIMAL BEHAVIOUR

Older but less wise

Most fish living in marine reserves are older, bigger and more fecund than those outside their borders, but they are also slower to flee a threat. The potential for 'spillover' of such fish into fisheries may boost support for reserves.

PETER F. SALE

Marine reserves are designed to protect the aquatic life within their margins. It seems probable that, by providing a place for fish to grow and reproduce, they will also benefit adjacent fisheries. Writing in *Ecology Letters*, Januchowski-Hartley *et al.*[1] add a new twist to this concept, by demonstrating that fish that wander over a reserve boundary into waters in which fishing occurs are more easily caught than those that have lived continuously with the threat of fishing.

Fish living in a reserve in which fishing is prohibited are expected to live longer, grow larger and have more offspring than those not so protected[2]. Some of these offspring will disperse naturally beyond the boundaries as eggs or larvae, in a process called recruitment subsidy; the remaining offspring generate greater fish numbers and biomass inside the reserve. This should, in turn, lead to 'spillover' — the continuous net movement of juvenile and adult individuals beyond reserve boundaries. The effects of recruitment subsidy are thought to be large and widespread because the larvae of most fish species are pelagic (they live within the water column, rather than near the bottom) and are therefore readily dispersed. By contrast, the spillover effect is expected to be restricted to sites close to the reserve boundary.

However, this eminently logical theory depends on assumptions that are seldom stated explicitly[2]. For this model to hold, the reserve must be large enough (or the fish sedentary enough) for individuals to experience reduced fishing pressure — a fish that wanders widely, swimming in and out of a reserve, will get little benefit. The reserve must also be well managed, so that fishing pressure within its boundaries really is lower than that outside. And the fish must produce offspring that are dispersive early in life. If these requirements are met, marine reserves should enhance fishery yields through both recruitment subsidy and spillover, and this expectation has become a commonly used argument in attempts to convince fishing communities to give up part of their fishing area for the establishment of a marine reserve.

The enthusiasm for marine reserves in the conservation community has led to strong advocacy for their use as a fishery-management tool — an advocacy effort that has jumped well ahead of the evidence that real augmentation of fishing actually results. Although the extent of this overreach has been noted several times[2,3], data showing that some reserves seem not to enhance fisheries[4] have been glossed over, and less-positive reports, such as one showing that the increase in marine protected areas worldwide is having no effect on the global loss of biodiversity[5], can have somewhat difficult paths to publication. Nevertheless, despite the hype, there is now abundant evidence that, for reasonably sedentary species in many environments, marine reserves will contain the more abundant and older fish expected by theory[2,4]. Concurrently, there is growing evidence of

REINHARD DIRSCHERL/FLPA

Figure 1 | Beyond the boundaries. Januchowski-Hartley *et al.*[1] compared the behaviour of fish species that are targets for fishing, including surgeonfish (such as *Naso vlamingii*, shown) and parrotfish, with non-targeted butterflyfish. They show that the flight initiation distance — the distance between a fish and an approaching snorkeller at which the fish flees — is smaller for targeted species in marine reserves than for those in fisheries. Moreover, this distance increases with increasing distance from the reserve boundary. No difference in flight behaviour was observed in the non-targeted species.

spillover, particularly in coral-reef systems[4,6], and some evidence for recruitment subsidy, in particular from a recent study of larval dispersal from marine reserves within the Great Barrier Reef Marine Park in Australia[7]. This study found that 55% (for stripey snapper; *Lutjanus carponotatus*) and 83% (for coral trout; *Plectropomus maculatus*) of all juveniles produced in reserves dispersed to subsidize recruitment to surrounding fishing grounds.

Now, Januchowski-Hartley and colleagues take the evidence for spillover a step further. They report that, in three Philippine locations, fish that live in the safety of a reserve are less vigilant than fish living outside the protected area, and that they bring their less-vigilant behaviour with them when they wander across a reserve boundary.

The researchers looked at the behaviour of three families of reef fish: butterflyfish (Chaetodontidae), which are not a target for fishing, and surgeonfish (Acanthuridae) and parrotfish (Scaridae), which are fishery targets (Fig. 1). The study design involved spotting individual fish engaged in feeding or swimming behaviour (not in social activities). A snorkeller (Januchowski-Hartley) would then descend to about 8–10 metres away from the fish, and swim towards it at constant slow speed. When the fish fled or sought shelter among the coral, the snorkeller recorded the linear distance between himself and the fish just before it fled — the flight initiation distance (FID). The authors also recorded the estimated total length of the fish. They classified each encounter as within one of eight 50-metre-wide zones that spanned from 200 m inside to 200 m outside the reserve boundary. Finally, they repeated the same procedure around control 'boundaries' within the fishing zone and away from marine reserves.

The main results are straightforward: the FID increased for each of the two fished families in all three reserves as the location of the encounters moved from inside to outside the boundary. No such trend was seen across control boundaries or for the non-fished butterflyfish. Because fish are easier to spear if they have a shorter FID, these results mean that fish just outside the boundary of a marine reserve will be easier to catch (at least by spearfishing) than those farther from the boundary.

The obvious explanation is that fish learn to adjust their FID according to the level of risk, that FIDs are shorter within reserves (where risk is low) and that fish that spill over into the fished area take time to increase their FIDs. These findings imply that the presence of a marine reserve is even more beneficial to a fishery than previously expected — not only do fish spill over, but also those that do are naive and more easily caught. Marine-reserve advocates will be delighted by this suggestion.

Are these results surprising, and are they important? Yes, and yes. I suspect that they will surprise those fishery scientists who use catch statistics as their primary data, and who tend to think of fish as hanging around waiting to be caught, although they will not surprise biologists who watch fish before the fish are caught. Fish are behaving beings, and are highly capable of adapting their behaviour to particular circumstances. Their behaviour as larvae, for example, gives them amazing control over their dispersal[8], and, as juveniles and adults, behaviour adds much of the complexity that builds up fish social structures and contributes to the challenges of managing some fisheries. Januchowski-Hartley and colleagues' results are a direct demonstration of an unexpected consequence of adaptive fish behaviour, and they will be relevant to anyone seeking to quantify the effects of a marine reserve on fishery yields. ■

Peter F. Sale *is at the United Nations University Institute for Water, Environment and Health, Hamilton, Ontario L8P 0A1, Canada.*
e-mail: sale@uwindsor.ca

1. Januchowski-Hartley, F. A., Graham, N. A. J., Cinner, J. E. & Russ, G. R. *Ecol. Lett.* http://dx.doi.org/10.1111/ele.12028 (2012).
2. Sale, P. F. *et al. Trends Ecol. Evol.* **20,** 74–80 (2005).
3. Willis, T. J., Millar, R. B., Babcock, R. C. & Tolimieri, N. *Environ. Conserv.* **30,** 97–103 (2003).
4. Russ, G. R. in *Coral Reef Fishes. Dynamics and Diversity in a Complex Ecosystem* (ed. Sale, P. F.) 421–443 (Academic, 2002).
5. Mora, C. & Sale, P. F. *Mar. Ecol. Prog. Ser.* **434,** 251–266 (2011).
6. McCook, L. J. *et al. Proc. Natl Acad. Sci. USA* **107,** 18278–18285 (2010).
7. Harrison, H. B. *et al. Curr. Biol.* **22,** 1023–1028 (2012).
8. Gerlach, G., Atema, J., Kingsford, M. J., Black, K. P. & Miller-Sims, V. *Proc. Natl Acad. Sci. USA* **104,** 858–863 (2007).

MATERIALS SCIENCE

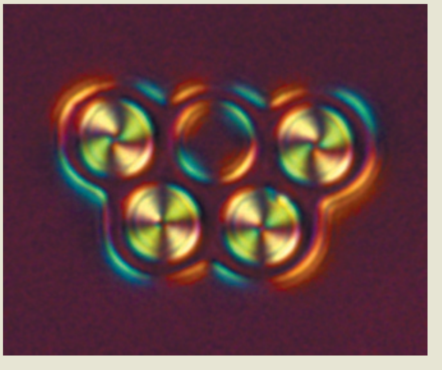
Topology matters

The effects of topology have been investigated in systems as diverse as molecules and the cosmos, but rarely at the micrometre scale. In this issue, Senyuk *et al*. explore how topology affects the alignment of micro-scale particles suspended in a 'nematic' host matrix — a liquid crystal in which the molecules are aligned but do not form well-defined planes (B. Senyuk *et al. Nature* **493,** 200–205; 2013).

The particles in naturally occurring colloids — systems of tiny particles of one material suspended in a different material — are often spheres or faceted crystals. Such particles are said to have a topology of genus (g) zero. To explore the effects of different topologies on colloid particles, Senyuk *et al*. synthesized particles constructed from rings 5–10 μm in diameter. These ranged from simple hoops ($g = 1$) to 'Olympic rings' ($g = 5$; pictured).

The authors found that when the particles were dispersed in a nematic liquid crystal, they typically aligned with their ring planes perpendicular to the direction of alignment of the liquid crystal (the 'director'). Each particle of genus g generated at least $g-1$ defects in the liquid-crystal matrix, in agreement with topological theorems. These were either point defects in the rings' holes, or loop defects that could run around the inside or outside of the rings. What's more, the orientation of the liquid-crystal molecules around each particle was dictated by the particle's topology.

Senyuk and co-workers went on to show that the particles could be aligned parallel to the director by melting and rapidly cooling the surrounding liquid crystals. In this case, defects appeared in the liquid crystal both within and next to the particles. And by applying an electric field at different rates, the authors could switch either the orientation of the colloidal particles or the configuration of the liquid-crystal molecules around the particles; the



resulting states were stable when the electric field was removed.

Finally, the researchers found that the particles diffused more easily along the nematic director than in other directions, and that the rate of diffusion decreased with increasing genus number. Taken together, these findings open up fresh opportunities for colloidal organization and self-assembly, as well as potential applications — in electro-optic or photonic devices, for example. **Rosamund Daw**

spillover, particularly in coral-reef systems[4,6], and some evidence for recruitment subsidy, in particular from a recent study of larval dispersal from marine reserves within the Great Barrier Reef Marine Park in Australia[7]. This study found that 55% (for stripey snapper; *Lutjanus carponotatus*) and 83% (for coral trout; *Plectropomus maculatus*) of all juveniles produced in reserves dispersed to subsidize recruitment to surrounding fishing grounds.

Now, Januchowski-Hartley and colleagues take the evidence for spillover a step further. They report that, in three Philippine locations, fish that live in the safety of a reserve are less vigilant than fish living outside the protected area, and that they bring their less-vigilant behaviour with them when they wander across a reserve boundary.

The researchers looked at the behaviour of three families of reef fish: butterflyfish (Chaetodontidae), which are not a target for fishing, and surgeonfish (Acanthuridae) and parrotfish (Scaridae), which are fishery targets (Fig. 1). The study design involved spotting individual fish engaged in feeding or swimming behaviour (not in social activities). A snorkeller (Januchowski-Hartley) would then descend to about 8–10 metres away from the fish, and swim towards it at constant slow speed. When the fish fled or sought shelter among the coral, the snorkeller recorded the linear distance between himself and the fish just before it fled — the flight initiation distance (FID). The authors also recorded the estimated total length of the fish. They classified each encounter as within one of eight 50-metre-wide zones that spanned from 200 m inside to 200 m outside the reserve boundary. Finally, they repeated the same procedure around control 'boundaries' within the fishing zone and away from marine reserves.

The main results are straightforward: the FID increased for each of the two fished families in all three reserves as the location of the encounters moved from inside to outside the boundary. No such trend was seen across control boundaries or for the non-fished butterflyfish. Because fish are easier to spear if they have a shorter FID, these results mean that fish just outside the boundary of a marine reserve will be easier to catch (at least by spearfishing) than those farther from the boundary.

The obvious explanation is that fish learn to adjust their FID according to the level of risk, that FIDs are shorter within reserves (where risk is low) and that fish that spill over into the fished area take time to increase their FIDs. These findings imply that the presence of a marine reserve is even more beneficial to a fishery than previously expected — not only do fish spill over, but also those that do are naive and more easily caught. Marine-reserve advocates will be delighted by this suggestion.

Are these results surprising, and are they important? Yes, and yes. I suspect that they will surprise those fishery scientists who use catch statistics as their primary data, and who tend to think of fish as hanging around waiting to be caught, although they will not surprise biologists who watch fish before the fish are caught. Fish are behaving beings, and are highly capable of adapting their behaviour to particular circumstances. Their behaviour as larvae, for example, gives them amazing control over their dispersal[8], and, as juveniles and adults, behaviour adds much of the complexity that builds up fish social structures and contributes to the challenges of managing some fisheries. Januchowski-Hartley and colleagues' results are a direct demonstration of an unexpected consequence of adaptive fish behaviour, and they will be relevant to anyone seeking to quantify the effects of a marine reserve on fishery yields. ■

Peter F. Sale *is at the United Nations University Institute for Water, Environment and Health, Hamilton, Ontario L8P 0A1, Canada.*
e-mail: sale@uwindsor.ca

1. Januchowski-Hartley, F. A., Graham, N. A. J., Cinner, J. E. & Russ, G. R. *Ecol. Lett.* http://dx.doi.org/10.1111/ele.12028 (2012).
2. Sale, P. F. *et al. Trends Ecol. Evol.* **20,** 74–80 (2005).
3. Willis, T. J., Millar, R. B., Babcock, R. C. & Tolimieri, N. *Environ. Conserv.* **30,** 97–103 (2003).
4. Russ, G. R. in *Coral Reef Fishes. Dynamics and Diversity in a Complex Ecosystem* (ed. Sale, P. F.) 421–443 (Academic, 2002).
5. Mora, C. & Sale, P. F. *Mar. Ecol. Prog. Ser.* **434,** 251–266 (2011).
6. McCook, L. J. *et al. Proc. Natl Acad. Sci. USA* **107,** 18278–18285 (2010).
7. Harrison, H. B. *et al. Curr. Biol.* **22,** 1023–1028 (2012).
8. Gerlach, G., Atema, J., Kingsford, M. J., Black, K. P. & Miller-Sims, V. *Proc. Natl Acad. Sci. USA* **104,** 858–863 (2007).

MATERIALS SCIENCE

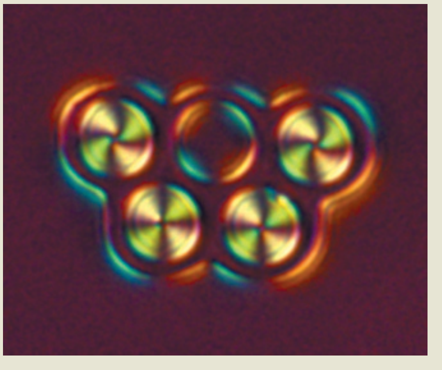
Topology matters

The effects of topology have been investigated in systems as diverse as molecules and the cosmos, but rarely at the micrometre scale. In this issue, Senyuk *et al*. explore how topology affects the alignment of micro-scale particles suspended in a 'nematic' host matrix — a liquid crystal in which the molecules are aligned but do not form well-defined planes (B. Senyuk *et al. Nature* **493,** 200–205; 2013).

The particles in naturally occurring colloids — systems of tiny particles of one material suspended in a different material — are often spheres or faceted crystals. Such particles are said to have a topology of genus (g) zero. To explore the effects of different topologies on colloid particles, Senyuk *et al*. synthesized particles constructed from rings 5–10 μm in diameter. These ranged from simple hoops ($g = 1$) to 'Olympic rings' ($g = 5$; pictured).

The authors found that when the particles were dispersed in a nematic liquid crystal, they typically aligned with their ring planes perpendicular to the direction of alignment of the liquid crystal (the 'director'). Each particle of genus g generated at least $g - 1$ defects in the liquid-crystal matrix, in agreement with topological theorems. These were either point defects in the rings' holes, or loop defects that could run around the inside or outside of the rings. What's more, the orientation of the liquid-crystal molecules around each particle was dictated by the particle's topology.

Senyuk and co-workers went on to show that the particles could be aligned parallel to the director by melting and rapidly cooling the surrounding liquid crystals. In this case, defects appeared in the liquid crystal both within and next to the particles. And by applying an electric field at different rates, the authors could switch either the orientation of the colloidal particles or the configuration of the liquid-crystal molecules around the particles; the



resulting states were stable when the electric field was removed.

Finally, the researchers found that the particles diffused more easily along the nematic director than in other directions, and that the rate of diffusion decreased with increasing genus number. Taken together, these findings open up fresh opportunities for colloidal organization and self-assembly, as well as potential applications — in electro-optic or photonic devices, for example. **Rosamund Daw**

Could one of these conditions be general — or 'ancestral' — for vertebrates, and the other have subsequently derived from it? Oisi *et al.* suggest that the cyclostome condition (a median NHP) is likely to be the earlier structure, because the monorhinal condition is found in the extinct jawless vertebrates commonly known as ostracoderms, which are between 480 million and 370 million years old and are regarded as more closely related to gnathostomes than to cyclostomes (Fig. 1).

Placodes are embryonic structures that are not preserved in ostracoderm fossils, but the skeletons of these animals provide information about their differing cranial anatomy[6,7]. Ostracoderms display diverse arrangements of the cranial spaces that housed the olfactory organs and the adenohypophysis, but in all of them these spaces open towards a median nostril, as in cyclostomes. For example, anaspids and osteostracans have a median dorsal nasohypophyseal complex similar to that of lampreys, which suggests some kind of cyclostome-like NHP, although other features show that osteostracans are closer to gnathostomes than to any other ostracoderm. By contrast, heterostracans and galeaspids had paired olfactory organs, still closely associated with the adenohypophysis in the latter, but opening anteriorly by a large median nostril and suggesting a broader NHP (Fig. 1).

This anatomical diversity is puzzling, but Oisi and colleagues' findings on hagfish development help to clarify the picture. The authors propose that the presence of the NHP in both hagfish and lamprey embryos supports the idea of a general, 'pan-cyclostome' pattern for craniofacial development, and that the anatomical characteristics that make present-day hagfish seem more primitive than lampreys are the result of a subsequent loss of features in hagfish during evolution. Such a pattern, combined with the fact that no NHP is seen during the development of any living gnathostomes, and the apparent progression from the monorhinal to diplorhinal condition in ostracoderms, suggests that a crucial innovation in the development of the vertebrate snout was the loss of a common nasohypophyseal duct and the rise of separate, paired nostrils. ■

Philippe Janvier *is at the Muséum National d'Histoire Naturelle, 75231 Paris Cedex 05, France.*
e-mail: janvier@mnhn.fr

1. Ota, K. G. & Kuratani, S. *Zool. Sci.* **23,** 403–418 (2006).
2. Oisi, Y., Ota, K. G., Kuraku, S., Fujimoto, S. & Kuratani, S. *Nature* **493,** 175–180 (2013).
3. Kuratani, S. *J. Anat.* **205,** 335–347 (2004).
4. Kuraku, S. *Zool. Sci.* **25,** 960–968 (2008).
5. Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. & Peterson, K. J. *Proc. Natl Acad. Sci. USA* **105,** 2946–2950 (2008).
6. Janvier, P. *Zool. Sci.* **25,** 1045–1056 (2008).
7. Gai, Z., Donoghue, P. C. J., Zhu, M., Janvier, P. & Stampanoni, M. *Nature* **476,** 324–327 (2011).

PHOTONICS

Phased array on a fingertip

An array of more than 4,000 optical antennas working in unison has been demonstrated on a millimetre-scale silicon chip. The result highlights the remarkable capabilities of optical integration in silicon. SEE LETTER P.195

THOMAS F. KRAUSS

In his Nobel lecture in 1909, physicist Ferdinand Braun remarked[1] that: "It had always seemed most desirable to me to transmit the waves, in the main, in one direction only". He was referring to radio waves, which had hitherto been known to radiate only in a doughnut shape away from the emitter, as described by Heinrich Hertz two decades earlier. The invention for which Braun was recognized was the phased array — the idea of connecting up multiple antennas fed from a common source and using the relationship between the antennas' phase (where a wave's peaks and troughs lie) to enhance the emission of radio waves in a given direction. His work was instrumental in the development of radio transmitters, and led to other inventions such as the television antenna and radar. The phased-array concept is now also extensively used in astronomy, where the information from multiple telescopes is collected in phase to extend the resolution of the system. On page 195 of this issue, Sun *et al.*[2] introduce an integrated optical phased array that combines the equivalent of more than 4,000 telescopes onto a single chip smaller than a fingernail.

The creation of phased arrays in the optical domain has been an active area of research for some time, as it is motivated by the need to improve applications such as optical beam steering and imaging. To achieve the highest performance, the challenges are to integrate a large number of antennas onto a single chip, to place them close together and to fully control the phase on each element. The short optical wavelength makes it easy, in principle, to address the first challenge and to build large arrays on a small footprint in an integrated format. The other two challenges are much more difficult to meet, however, because the optical couplers that feed a controlled amount of power into each antenna and the phase shifters needed to control the phase on each element tend to be tens or hundreds of wavelengths long, which requires a large distance between each antenna. Therefore, the latest realizations of integrated optical phased arrays have been relatively large and have featured a maximum of only 4×4 antennas[3,4].

The architecture now introduced by Sun *et al.*[2] addresses all of these requirements and is scalable. By carefully designing the optical coupler, phase shifter and gratings that act as radiating antennas, the team has managed to produce a phased array of 64×64 antenna units, or 'pixels', on a footprint of a little more than 0.576×0.576 millimetres, with each pixel covering 9×9 micrometres — equivalent to a pixel side length of six optical wavelengths. Most impressive are the accuracy and control that the researchers were able to exercise over the array's optical functions, especially those of the optical couplers.

The device's accuracy and performance are highlighted by the well-balanced power output achieved across the array. Equally, the fact that a phase shift of π radians can be attained on a pixel size a few wavelengths across (in contrast to the tens or hundreds of wavelengths previously required), and with relatively low power and crosstalk, points to the remarkable capabilities of nanophotonic concepts that are now available. Fabrication of the array, as a photonic circuit, using only silicon microelectronics processing techniques — a concept pioneered[5] by the IMEC microelectronics centre in Leuven, Belgium — is also a testimony to the high performance that can be achieved by using silicon as a photonic material. The successful demonstration of the authors' array shows that the large-scale integration of photonic circuitry in silicon has well and truly arrived.

So where do we go from here? One interesting idea is that of imaging through light-scattering media[6], such as roughened glass or biological tissue. Such imaging uses the adaptive-optics principle of adjusting the phase of an optical wave locally so as to compensate for the distortion caused by the medium. The authors' optical phased array now makes available the phase control and large pixel number that this application requires. To achieve imaging through time-varying light scatterers, such as turbulent liquids, a fast time-response of the system is also essential; the small and efficient tuning of each antenna demonstrated by Sun *et al.* can also meet this requirement. Another application is in sensor networks, in which communication between randomly distributed sensors is based on

optical beams. Such networks require accurate control over the direction of each individual beam, which can now be achieved with the authors' array.

Other applications in imaging, such as projection television, involve visible optical wavelengths that are shorter than the near-infrared wavelengths used by the authors. Operation at these wavelengths would require further miniaturization of the array's pixel size and waveguides that are not made of silicon, because silicon absorbs light in the visible range. Also, the multiple images seen in the device's radiation pattern (see Fig. 2 of the paper[2]) are due to the fact that the pixels are multiple wavelengths apart. To reduce the number of images, ideally to a single one, it will be necessary to make the pixels even smaller and to move them closer together. This might be achievable by stacking the optical coupler, phase shifter and grating antenna on top of one another, thereby integrating the array in three dimensions. Clearly, further miniaturization and three-dimensional integration are exciting challenges that can now be considered seriously.

Most remarkably, even a century ago Braun had already thought ahead, and considered whether the antenna effects he described might also be observable in the optical domain. He wondered[1] whether the concepts he had studied with radio waves could "also be linked up with optical phenomena, though this can hardly be experimentally verified in this field".

Now, Sun *et al.* have provided us with just such an experimental demonstration. Braun would surely have been pleased indeed. ■

Thomas F. Krauss *is in the Department of Physics, University of York, York YO10 5DD, UK.*
e-mail: thomas.krauss@york.ac.uk

1. Braun, K. F. http://nobelprize.org/nobel_prizes/physics/laureates/1909/braun-lecture.pdf 239–241 (1909).
2. Sun, J. *et al. Nature* **493,** 195–199 (2013).
3. Van Acoleyen, K., Rogier, H. & Baets, R. *Opt. Express* **18,** 13655–13660 (2010).
4. Doylend, J. K. *et al. Opt. Express* **19,** 21595–21604 (2011).
5. www.epixfab.eu
6. Čižmár, T., Mazilu, M. & Dholakia, K. *Nature Photon.* **4,** 388–394 (2010).

STRUCTURAL BIOLOGY

Insulin meets its receptor

The hormone insulin has a central role in human physiology, yet the answer to a fundamental biochemical question — how it binds to its cell-surface receptor — has remained elusive, until now. SEE LETTER P.241

STEVAN R. HUBBARD

Insulin holds a storied place in the fields of physiology and biochemistry[1]. First isolated in 1921, it was soon used to save the lives of individuals with diabetes. This medical breakthrough and subsequent biochemical research on insulin have garnered three Nobel prizes. The three-dimensional structure of insulin was determined by X-ray crystallography in 1969. Yet, despite intense interest, an atomic view of insulin bound to its receptor has been lacking — a situation rectified by the report of Menting *et al.*[2] in this issue.

Over the years, researchers have used protein crystallography to determine the structure of numerous hormones, growth factors and cytokines bound to their receptors[3,4]. In general, structures of these ligand–receptor systems were tractable because the ligand-binding domains of the receptors were of moderate size and many could be produced in bacteria. Moreover, the binding modes of these ligands to their receptors are relatively simple, involving one or two receptor subdomains that are contiguous in amino-acid sequence. However, the production of insulin-receptor protein for structural studies is fraught with challenges.

The insulin receptor belongs to the receptor tyrosine-kinase family of cell-surface receptors[5]. These are single-pass transmembrane proteins with an extracellular region that binds a protein ligand (typically a growth factor) and a cytoplasmic region that contains a tyrosine-kinase domain, which phosphorylates specific tyrosine amino-acid residues on other signalling proteins, as well as on the receptor itself. On ligand binding, many receptor tyrosine kinases form a dimer, which facilitates tyrosine phosphorylation of one cytoplasmic domain by the other, thereby activating the receptor[6]. By contrast, the insulin receptor is a preformed, disulphide-bridged dimer,

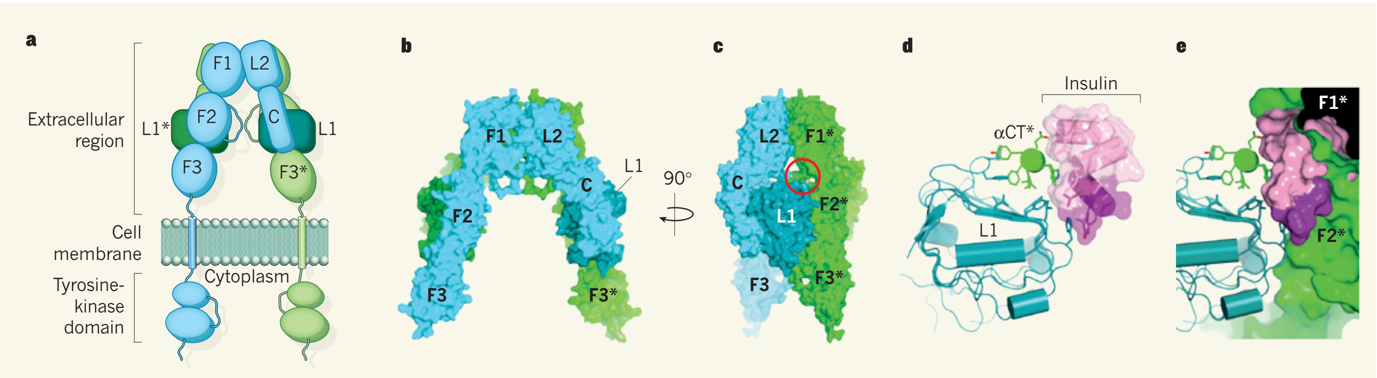


Figure 1 | Binding of insulin to its receptor. a, Diagram of the dimeric insulin receptor. The extracellular region comprises domains L1, C, L2, F1, F2 and F3. One half-receptor (at the back) is coloured green, with its domain labels marked by an asterisk, and the other half-receptor is in blue; L1 and L1* are coloured in darker shades. The carboxy-terminal region of the α-chain (αCT) extends from the first half of F2 and folds back onto L1 from the opposite half-receptor. **b**, **c**, Crystal structure of the extracellular region of the *apo* insulin receptor[8], shown in surface representation (**b**) and rotated by 90° (**c**). The colouring is consistent with that in **a**. The red circle points to one of the two equivalent insulin-binding sites. **d**, Menting *et al.*[2] present the structure of insulin bound to L1 and αCT. The viewing angle of L1 is approximately the same as in **c**. Side chains of interacting residues in insulin, L1 and αCT are shown in stick representation. Residues in the A and B chains of insulin are coloured pink and purple, respectively. **e**, As in **d**, but with the other half-receptor superimposed from the structure of the *apo* receptor (as in **c**).

optical beams. Such networks require accurate control over the direction of each individual beam, which can now be achieved with the authors' array.

Other applications in imaging, such as projection television, involve visible optical wavelengths that are shorter than the near-infrared wavelengths used by the authors. Operation at these wavelengths would require further miniaturization of the array's pixel size and waveguides that are not made of silicon, because silicon absorbs light in the visible range. Also, the multiple images seen in the device's radiation pattern (see Fig. 2 of the paper[2]) are due to the fact that the pixels are multiple wavelengths apart. To reduce the number of images, ideally to a single one, it will be necessary to make the pixels even smaller and to move them closer together. This might be achievable by stacking the optical coupler, phase shifter and grating antenna on top of one another, thereby integrating the array in three dimensions. Clearly, further miniaturization and three-dimensional integration are exciting challenges that can now be considered seriously.

Most remarkably, even a century ago Braun had already thought ahead, and considered whether the antenna effects he described might also be observable in the optical domain. He wondered[1] whether the concepts he had studied with radio waves could "also be linked up with optical phenomena, though this can hardly be experimentally verified in this field". Now, Sun *et al.* have provided us with just such an experimental demonstration. Braun would surely have been pleased indeed. ■

Thomas F. Krauss *is in the Department of Physics, University of York, York YO10 5DD, UK.*
e-mail: thomas.krauss@york.ac.uk

1. Braun, K. F. http://nobelprize.org/nobel_prizes/physics/laureates/1909/braun-lecture.pdf 239–241 (1909).
2. Sun, J. *et al. Nature* **493,** 195–199 (2013).
3. Van Acoleyen, K., Rogier, H. & Baets, R. *Opt. Express* **18,** 13655–13660 (2010).
4. Doylend, J. K. *et al. Opt. Express* **19,** 21595–21604 (2011).
5. www.epixfab.eu
6. Čižmár, T., Mazilu, M. & Dholakia, K. *Nature Photon.* **4,** 388–394 (2010).

STRUCTURAL BIOLOGY

Insulin meets its receptor

The hormone insulin has a central role in human physiology, yet the answer to a fundamental biochemical question — how it binds to its cell-surface receptor — has remained elusive, until now. See Letter p.241

STEVAN R. HUBBARD

Insulin holds a storied place in the fields of physiology and biochemistry[1]. First isolated in 1921, it was soon used to save the lives of individuals with diabetes. This medical breakthrough and subsequent biochemical research on insulin have garnered three Nobel prizes. The three-dimensional structure of insulin was determined by X-ray crystallography in 1969. Yet, despite intense interest, an atomic view of insulin bound to its receptor has been lacking — a situation rectified by the report of Menting *et al.*[2] in this issue.

Over the years, researchers have used protein crystallography to determine the structure of numerous hormones, growth factors and cytokines bound to their receptors[3,4]. In general, structures of these ligand–receptor systems were tractable because the ligand-binding domains of the receptors were of moderate size and many could be produced in bacteria. Moreover, the binding modes of these ligands to their receptors are relatively simple, involving one or two receptor subdomains that are contiguous in amino-acid sequence. However, the production of insulin-receptor protein for structural studies is fraught with challenges.

The insulin receptor belongs to the receptor tyrosine-kinase family of cell-surface receptors[5]. These are single-pass transmembrane proteins with an extracellular region that binds a protein ligand (typically a growth factor) and a cytoplasmic region that contains a tyrosine-kinase domain, which phosphorylates specific tyrosine amino-acid residues on other signalling proteins, as well as on the receptor itself. On ligand binding, many receptor tyrosine kinases form a dimer, which facilitates tyrosine phosphorylation of one cytoplasmic domain by the other, thereby activating the receptor[6]. By contrast, the insulin receptor is a preformed, disulphide-bridged dimer,

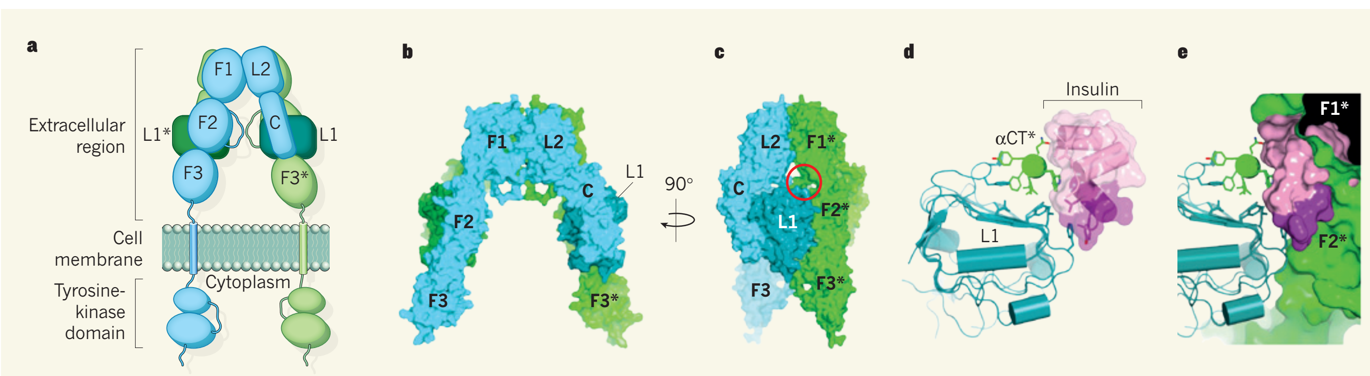


Figure 1 | Binding of insulin to its receptor. a, Diagram of the dimeric insulin receptor. The extracellular region comprises domains L1, C, L2, F1, F2 and F3. One half-receptor (at the back) is coloured green, with its domain labels marked by an asterisk, and the other half-receptor is in blue; L1 and L1* are coloured in darker shades. The carboxy-terminal region of the α-chain (αCT) extends from the first half of F2 and folds back onto L1 from the opposite half-receptor. **b**, **c**, Crystal structure of the extracellular region of the *apo* insulin receptor[8], shown in surface representation (**b**) and rotated by 90° (**c**). The colouring is consistent with that in **a**. The red circle points to one of the two equivalent insulin-binding sites. **d**, Menting *et al.*[2] present the structure of insulin bound to L1 and αCT. The viewing angle of L1 is approximately the same as in **c**. Side chains of interacting residues in insulin, L1 and αCT are shown in stick representation. Residues in the A and B chains of insulin are coloured pink and purple, respectively. **e**, As in **d**, but with the other half-receptor superimposed from the structure of the *apo* receptor (as in **c**).

and insulin binding is thought to induce a conformational change in the receptor that initiates cross-phosphorylation of the two cytoplasmic domains.

The mature insulin receptor consists of two copies of two polypeptide chains, α and β. The α-chain (723 amino-acid residues) is entirely extracellular and heavily glycosylated. The β-chain (620 residues) starts on the extracellular side and spans the membrane (by way of an α-helix) to the cytoplasmic side, which harbours the tyrosine-kinase domain (Fig. 1a). Each α-chain is linked to a β-chain through a disulphide bridge to form an αβ half-receptor, and the two half-receptors are linked by at least two (and possibly four) disulphide bridges. The extracellular region of each half-receptor contains a series of folded domains: L1, C, L2, F1, F2 and F3 (Fig. 1a).

Mature insulin consists of two polypeptide chains: an A chain of 21 residues and a B chain of 30 residues. It contains one intrachain disulphide bridge (in the A chain) and two interchain disulphide bridges.

Previous biochemical studies[1,7] established that a single insulin molecule binds, with sub-nanomolar affinity, to the dimeric receptor (1:2 stoichiometry) and activates it. The two equivalent insulin-binding sites on the receptor are composed of two distinct interfaces. The primary interface consists of the L1 domain of one half-receptor and the carboxy-terminal region of the α-chain (αCT) of the second half-receptor. The secondary interface comprises the loop regions near the junction of F1 and F2 in the second half-receptor.

In 2006, some of the authors of the current study determined the atomic structure of the dimeric extracellular portion of the insulin receptor in the *apo* conformation (devoid of insulin). This landmark structure[8] revealed an inverted V shape, with the two half-receptors arranged in an antiparallel fashion (Fig. 1b,c). Although an insulin-mimetic peptide was present in the crystallization trials, it could not be located in the electron-density map and was presumed to be unbound.

Menting *et al.* now use some of the same strategies to crystallize truncated versions of the α-chain, containing only the primary binding interface (involving L1 and αCT), in the presence of insulin. Because the αCT region is a considerable distance in sequence from L1, they either added it as a synthetic peptide or spliced it onto the C-terminal end of the truncated α-chain construct. This time, the catch was made: insulin was visible in the electron-density maps along with the receptor.

In the four crystal structures the authors determined, insulin is bound to the L1 domain, but just barely (Fig. 1d). Almost all of the hydrophobic residues extending from the flat β-sheet surface of L1, which were previously mapped[1] as insulin-binding 'hotspots', contact the αCT helix rather than insulin, and it is the αCT helix that is in intimate contact with insulin.

Two structural transitions occur on insulin binding, one in insulin and one in αCT (L1 is not appreciably altered). In structures of insulin alone, residues at the C-terminal end of the B chain pack against the rest of the molecule, but are displaced on binding to αCT. This observation confirms a long-standing prediction that such a structural transition occurs in insulin on binding to its receptor[1]. What was not predicted, however, was the behaviour of αCT. In the *apo* structure, the αCT helix is bound to L1, but insulin binding causes a repositioning of the helix on the L1 surface such that it interacts with insulin, as well as with L1.

Menting and colleagues' structures also provide insights into the receptor-activation mechanism. If the L1 domain from the insulin-bound structure is superimposed on L1 from the *apo* structure, it is apparent that a conformational change must take place in the other half-receptor (near F1 and F2, the secondary binding interface) to accommodate insulin binding to αCT–L1 (Fig. 1e). Evidently, this change underlies the structural mechanism whereby insulin binding triggers cross-phosphorylation of the cytoplasmic kinase domains of its receptor. The precise nature of this conformational change and how it repositions the kinase domains remain unresolved.

Fast-acting insulin analogues have been used to treat patients with diabetes for many years, and the three-dimensional structure of insulin was instrumental in their design[9]. The structure of insulin bound to its receptor is not only the latest milestone in insulin research, but should also afford opportunities for the design of insulin analogues with enhanced receptor binding, as well as favourable pharmacokinetics. ■

Stevan R. Hubbard *is in the Structural Biology Program, Kimmel Center for Biology and Medicine at the Skirball Institute of Biomolecular Medicine, Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA.*
e-mail: stevan.hubbard@med.nyu.edu

1. Ward, C. W. & Lawrence, M. C. *Front. Endocrinol.* **2,** 76 (2011).
2. Menting, J. G. *et al. Nature* **493,** 241–245 (2013).
3. Stroud, R. M. & Wells, J. A. *Sci. STKE* **2004,** re7 (2004).
4. Wang, X., Lupardus, P., Laporte, S. L. & Garcia, K. C. *Annu. Rev. Immunol.* **27,** 29–60 (2009).
5. Lemmon, M. A. & Schlessinger, J. *Cell* **141,** 1117–1134 (2010).
6. Hubbard, S. R. & Miller, W. T. *Curr. Opin. Cell Biol.* **19,** 117–123 (2007).
7. De Meyts, P. *Trends Biochem. Sci.* **33,** 376–384 (2008).
8. McKern, N. M. *et al. Nature* **443,** 218–221 (2006).
9. Pandyarajan, V. & Weiss, M. A. *Curr. Diab. Rep.* **12,** 697–704 (2012).

POLYMER CHEMISTRY

Wasted loops quantified

A method for dissecting the polymeric networks of gels enables the number of loops — strands that connect to themselves — within them to be counted. This allows network morphologies to be correlated with gel properties.

ANNA C. BALAZS

Products as diverse as jelly, pharmaceuticals and cosmetics often contain a common ingredient: a gel formed from an extensive network of cross-linked polymer chains. Such gels typically encase a large amount of fluid. Hydrogels, for example, contain as much as 90% water, yet can still be as strong and tough as rubber[1]. The strength of these materials comes from the polymers that interconnect multiple chains within the network, whereas strands that form non-connecting loops weaken the material (Fig. 1). Writing in *Proceedings of the National Academy of Sciences*, Zhou *et al.*[2] report an experimental method for determining the number of loops — something that wasn't previously possible. The ability to quantify the fraction of a gel that consists of such 'wasted' loops[3] should enable researchers to establish synthetic routes that minimize the formation of these topologies, and thereby optimize the strength and toughness of such materials.

Zhou *et al.* cleverly realized that to determine the fraction of loops in a polymer gel, they had to dissect it, and so they designed a material that could be carefully and precisely broken apart. Their hydrogel was composed of a linear polymer that contained a reactive group for forming cross-links at each end, and a branched monomer that contained three reactive end groups. The linear polymer was designed to contain a strategically placed, readily cleavable chemical bond.

The researchers made the gel by combining the components in solution, and then applied their careful dissection technique to break the readily cleavable bonds. Because these bonds were precisely placed, only four products would form from this degradation reaction if an ideal network without any loops had been made; each of those products had its own distinctive molecular mass. But if some

and insulin binding is thought to induce a conformational change in the receptor that initiates cross-phosphorylation of the two cytoplasmic domains.

The mature insulin receptor consists of two copies of two polypeptide chains, α and β. The α-chain (723 amino-acid residues) is entirely extracellular and heavily glycosylated. The β-chain (620 residues) starts on the extracellular side and spans the membrane (by way of an α-helix) to the cytoplasmic side, which harbours the tyrosine-kinase domain (Fig. 1a). Each α-chain is linked to a β-chain through a disulphide bridge to form an αβ half-receptor, and the two half-receptors are linked by at least two (and possibly four) disulphide bridges. The extracellular region of each half-receptor contains a series of folded domains: L1, C, L2, F1, F2 and F3 (Fig. 1a).

Mature insulin consists of two polypeptide chains: an A chain of 21 residues and a B chain of 30 residues. It contains one intrachain disulphide bridge (in the A chain) and two interchain disulphide bridges.

Previous biochemical studies[1,7] established that a single insulin molecule binds, with sub-nanomolar affinity, to the dimeric receptor (1:2 stoichiometry) and activates it. The two equivalent insulin-binding sites on the receptor are composed of two distinct interfaces. The primary interface consists of the L1 domain of one half-receptor and the carboxy-terminal region of the α-chain (αCT) of the second half-receptor. The secondary interface comprises the loop regions near the junction of F1 and F2 in the second half-receptor.

In 2006, some of the authors of the current study determined the atomic structure of the dimeric extracellular portion of the insulin receptor in the *apo* conformation (devoid of insulin). This landmark structure[8] revealed an inverted V shape, with the two half-receptors arranged in an antiparallel fashion (Fig. 1b,c). Although an insulin-mimetic peptide was present in the crystallization trials, it could not be located in the electron-density map and was presumed to be unbound.

Menting *et al.* now use some of the same strategies to crystallize truncated versions of the α-chain, containing only the primary binding interface (involving L1 and αCT), in the presence of insulin. Because the αCT region is a considerable distance in sequence from L1, they either added it as a synthetic peptide or spliced it onto the C-terminal end of the truncated α-chain construct. This time, the catch was made: insulin was visible in the electron-density maps along with the receptor.

In the four crystal structures the authors determined, insulin is bound to the L1 domain, but just barely (Fig. 1d). Almost all of the hydrophobic residues extending from the flat β-sheet surface of L1, which were previously mapped[1] as insulin-binding 'hotspots', contact the αCT helix rather than insulin, and it is the αCT helix that is in intimate contact with insulin.

Two structural transitions occur on insulin binding, one in insulin and one in αCT (L1 is not appreciably altered). In structures of insulin alone, residues at the C-terminal end of the B chain pack against the rest of the molecule, but are displaced on binding to αCT. This observation confirms a long-standing prediction that such a structural transition occurs in insulin on binding to its receptor[1]. What was not predicted, however, was the behaviour of αCT. In the *apo* structure, the αCT helix is bound to L1, but insulin binding causes a repositioning of the helix on the L1 surface such that it interacts with insulin, as well as with L1.

Menting and colleagues' structures also provide insights into the receptor-activation mechanism. If the L1 domain from the insulin-bound structure is superimposed on L1 from the *apo* structure, it is apparent that a conformational change must take place in the other half-receptor (near F1 and F2, the secondary binding interface) to accommodate insulin binding to αCT–L1 (Fig. 1e). Evidently, this change underlies the structural mechanism whereby insulin binding triggers cross-phosphorylation of the cytoplasmic kinase domains of its receptor. The precise nature of this conformational change and how it repositions the kinase domains remain unresolved.

Fast-acting insulin analogues have been used to treat patients with diabetes for many years, and the three-dimensional structure of insulin was instrumental in their design[9]. The structure of insulin bound to its receptor is not only the latest milestone in insulin research, but should also afford opportunities for the design of insulin analogues with enhanced receptor binding, as well as favourable pharmacokinetics. ■

Stevan R. Hubbard *is in the Structural Biology Program, Kimmel Center for Biology and Medicine at the Skirball Institute of Biomolecular Medicine, Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA.*
e-mail: stevan.hubbard@med.nyu.edu

1. Ward, C. W. & Lawrence, M. C. *Front. Endocrinol.* **2,** 76 (2011).
2. Menting, J. G. *et al. Nature* **493,** 241–245 (2013).
3. Stroud, R. M. & Wells, J. A. *Sci. STKE* **2004,** re7 (2004).
4. Wang, X., Lupardus, P., Laporte, S. L. & Garcia, K. C. *Annu. Rev. Immunol.* **27,** 29–60 (2009).
5. Lemmon, M. A. & Schlessinger, J. *Cell* **141,** 1117–1134 (2010).
6. Hubbard, S. R. & Miller, W. T. *Curr. Opin. Cell Biol.* **19,** 117–123 (2007).
7. De Meyts, P. *Trends Biochem. Sci.* **33,** 376–384 (2008).
8. McKern, N. M. *et al. Nature* **443,** 218–221 (2006).
9. Pandyarajan, V. & Weiss, M. A. *Curr. Diab. Rep.* **12,** 697–704 (2012).

POLYMER CHEMISTRY

Wasted loops quantified

A method for dissecting the polymeric networks of gels enables the number of loops — strands that connect to themselves — within them to be counted. This allows network morphologies to be correlated with gel properties.

ANNA C. BALAZS

Products as diverse as jelly, pharmaceuticals and cosmetics often contain a common ingredient: a gel formed from an extensive network of cross-linked polymer chains. Such gels typically encase a large amount of fluid. Hydrogels, for example, contain as much as 90% water, yet can still be as strong and tough as rubber[1]. The strength of these materials comes from the polymers that interconnect multiple chains within the network, whereas strands that form non-connecting loops weaken the material (Fig. 1). Writing in *Proceedings of the National Academy of Sciences*, Zhou *et al.*[2] report an experimental method for determining the number of loops — something that wasn't previously possible. The ability to quantify the fraction of a gel that consists of such 'wasted' loops[3] should enable researchers to establish synthetic routes that minimize the formation of these topologies, and thereby optimize the strength and toughness of such materials.

Zhou *et al.* cleverly realized that to determine the fraction of loops in a polymer gel, they had to dissect it, and so they designed a material that could be carefully and precisely broken apart. Their hydrogel was composed of a linear polymer that contained a reactive group for forming cross-links at each end, and a branched monomer that contained three reactive end groups. The linear polymer was designed to contain a strategically placed, readily cleavable chemical bond.

The researchers made the gel by combining the components in solution, and then applied their careful dissection technique to break the readily cleavable bonds. Because these bonds were precisely placed, only four products would form from this degradation reaction if an ideal network without any loops had been made; each of those products had its own distinctive molecular mass. But if some

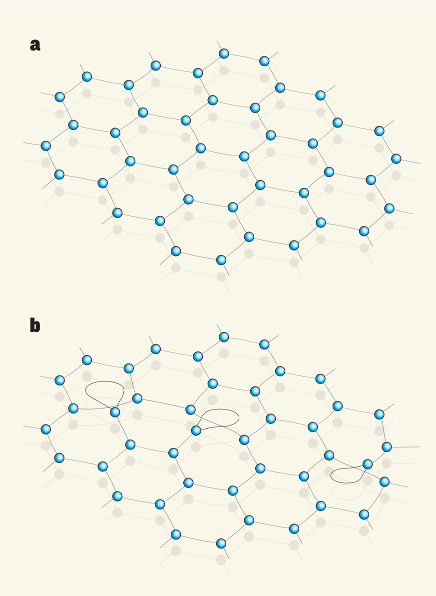


Figure 1 | Primary loops in cross-linked polymer networks. **a**, In an ideal polymer network, every polymer strand forms connections to other strands, strengthening the resulting material. **b**, Primary loops form if a strand closes on itself without connecting to other chains. The presence of such loops reduces the strength of the material. Zhou *et al.*[2] report a method for quantifying the number of loops in the polymer network of a gel.

loops had formed, different degradation products (with their own characteristic molecular masses) would be obtained. By measuring the masses of the actual degradation products and comparing them to the predicted masses for the ideal network, the authors could therefore determine the number of loops in the system. They named this technique 'network disassembly spectrometry' (NDS).

A potential limitation of NDS is that one must be able to predict the degradation products in order to apply the technique. Nevertheless, a broad range of polymers and chemistries will be amenable to this analytical approach. What's more, the technique enables new fundamental studies to be made of the dynamics of polymer-network formation. For example, by using NDS at different stages of their network's formation, Zhou *et al.* were able to probe the structural evolution of the gel. And by varying the reaction conditions of network assembly, they could control the fraction of loops that formed in the system, and thus gain valuable insight into the factors that contribute to the formation of these 'defects'.

One application of NDS might be as a tool for correlating the microstructure of gels with their mechanical properties. This could be useful in the development of artificial muscles that perform useful work[4]. For example, some gels controllably expand and contract in response to external stimuli (such as changes in pH, illumination or heat)[5] or internal chemical reactions[4], and so can function as actuators that drive movement within a system.

The rhythmic swelling and deswelling of gels has been harnessed to push micrometre-sized particles and cells along a surface[6]. Heat-responsive gels have also been used to drive micrometre-sized posts in and out of a bath of solution[7]: when the system was hot, the gel shrank, pulling the posts out of the bath, but when the system cooled down, the gel expanded and pushed the posts back in. This behaviour formed the basis of a homeostatic device that autonomously regulated the temperature of the system. For optimal performance, such soft actuators need to be mechanically robust — which means that they should not contain wasted loops.

On the other hand, because NDS can yield correlations between mechanical properties and the fraction of loops in a gel, it might help to provide design rules for exploiting loops, perhaps leading to the development of robust materials that nonetheless are highly flexible. In particular, there is rapidly growing interest in reconfigurable materials that can dramatically change shape in response to external cues[8–10]. Examples abound in biology, because such adaptive behaviour is vital for survival — consider, for example, the ability of octopuses and cuttlefish to change their shape, colour and texture in order to camouflage themselves in the presence of predators. Because gels can be made to shrink and swell controllably, they can be driven to change shape, and thus are ideal synthetic materials for creating systems that accomplish analogous adaptive behaviours.

Advances in this area could lead to multi-functional systems that have one structure and function in one environment, but another shape and function under different conditions. However, to undergo significant shape changes, the material must be very flexible. It might be that the introduction of just the right amount of loops within a polymer network could lead to robust, but sufficiently flexible, shape-changing materials. With findings from NDS, perhaps researchers will eventually establish effective routes for making gels that act like cuttlefish. ■

Anna C. Balazs *is in the Chemical Engineering Department, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA.*
e-mail: balazs@pitt.edu

1. Gong, J. P. *Soft Matter* **6,** 2583–2590 (2010).
2. Zhou, H. *et al. Proc. Natl Acad. Sci. USA* **109,** 19119–19124 (2012).
3. Panyukov, S. & Rabin, Y. *Phys. Rep.* **269,** 1–131 (1996).
4. Yashin, V., Kuksenok, O., Dayal, P. & Balazs, A. C. *Rep. Progr. Phys.* **75,** 066601 (2012).
5. Cohen Stuart, M. A. *et al. Nature Mater.* **9,** 101–113 (2010).
6. Yoshida, R. *Sensors* **10,** 1810–1822 (2010).
7. He, X. *et al. Nature* **487,** 214–218 (2012).
8. Guillet, P. *et al. Soft Matter* **5,** 3409–3411 (2009).
9. Yashin, V. V., Kuksenok, O. & Balazs, A. C. *J. Phys. Chem. B* **114,** 6316–6322 (2010).
10. Ueno, T., Bundo, K., Akagi, Y., Sakai, T. & Yoshida, R. *Soft Matter* **6,** 6072–6074 (2010).

EARTH SCIENCE

How glaciers grow

A state-of-the-art numerical model shows that the advance of glaciers in a cooling climate depends strongly on the pre-existing landscape, and that glacial erosion paves the way for greater glacial extent in the future. **See Letter p.206**

SIMON H. BROCKLEHURST

The growth of glaciers reflects the balance between the accumulation of snow and its loss through melting. Given the close association between altitude and temperature[1], the elevation and morphology of the valley floor on which a glacier forms are key determinants of a glacier's size and longevity. Glacial erosion has carved the spectacular alpine landscapes of many mountain ranges, and these are characterized by extensive glacial valley floors at elevations close to the long-term snowline[2]. These landscapes contrast strongly with their precursors, which are generally steeper, narrower valleys sculpted by rivers (Fig. 1). On page 206 of this issue, Pedersen and Egholm[3] use a state-of-the-art numerical model of glacier dynamics to quantify, for the first time, the stark contrast in glacier development between modern alpine mountain ranges and those that came before them. They compare the legacy of numerous glaciations during the Pleistocene epoch (between about 2.5 million and 10,000 years ago) with the landscapes that would have been present at the onset of these glaciations.

It has long been recognized that glaciers and their surroundings share an intimate, coupled relationship[4,5]. Shaded aspects and large areas at high elevation promote glacier growth. In turn, glacial erosion strongly modifies the underlying landscape, widening valley floors and making hill slopes steeper. However, given the efficiency of glaciers at reworking and removing the evidence left by previous glacial cycles, direct insight into the relationship between glaciers and topography during

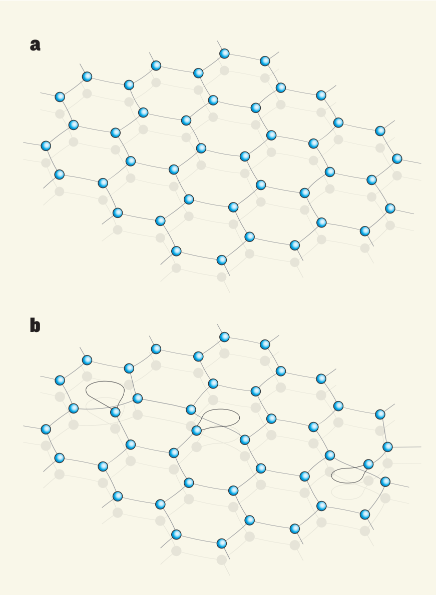


Figure 1 | Primary loops in cross-linked polymer networks. **a**, In an ideal polymer network, every polymer strand forms connections to other strands, strengthening the resulting material. **b**, Primary loops form if a strand closes on itself without connecting to other chains. The presence of such loops reduces the strength of the material. Zhou *et al.*[2] report a method for quantifying the number of loops in the polymer network of a gel.

loops had formed, different degradation products (with their own characteristic molecular masses) would be obtained. By measuring the masses of the actual degradation products and comparing them to the predicted masses for the ideal network, the authors could therefore determine the number of loops in the system. They named this technique 'network disassembly spectrometry' (NDS).

A potential limitation of NDS is that one must be able to predict the degradation products in order to apply the technique. Nevertheless, a broad range of polymers and chemistries will be amenable to this analytical approach. What's more, the technique enables new fundamental studies to be made of the dynamics of polymer-network formation. For example, by using NDS at different stages of their network's formation, Zhou *et al.* were able to probe the structural evolution of the gel. And by varying the reaction conditions of network assembly, they could control the fraction of loops that formed in the system, and thus gain valuable insight into the factors that contribute to the formation of these 'defects'.

One application of NDS might be as a tool for correlating the microstructure of gels with their mechanical properties. This could be useful in the development of artificial muscles that perform useful work[4]. For example, some gels controllably expand and contract in response to external stimuli (such as changes in pH, illumination or heat)[5] or internal chemical reactions[4], and so can function as actuators that drive movement within a system.

The rhythmic swelling and deswelling of gels has been harnessed to push micrometre-sized particles and cells along a surface[6]. Heat-responsive gels have also been used to drive micrometre-sized posts in and out of a bath of solution[7]: when the system was hot, the gel shrank, pulling the posts out of the bath, but when the system cooled down, the gel expanded and pushed the posts back in. This behaviour formed the basis of a homeostatic device that autonomously regulated the temperature of the system. For optimal performance, such soft actuators need to be mechanically robust — which means that they should not contain wasted loops.

On the other hand, because NDS can yield correlations between mechanical properties and the fraction of loops in a gel, it might help to provide design rules for exploiting loops, perhaps leading to the development of robust materials that nonetheless are highly flexible. In particular, there is rapidly growing interest in reconfigurable materials that can dramatically change shape in response to external cues[8–10]. Examples abound in biology, because such adaptive behaviour is vital for survival — consider, for example, the ability of octopuses and cuttlefish to change their shape, colour and texture in order to camouflage themselves in the presence of predators. Because gels can be made to shrink and swell controllably, they can be driven to change shape, and thus are ideal synthetic materials for creating systems that accomplish analogous adaptive behaviours.

Advances in this area could lead to multifunctional systems that have one structure and function in one environment, but another shape and function under different conditions. However, to undergo significant shape changes, the material must be very flexible. It might be that the introduction of just the right amount of loops within a polymer network could lead to robust, but sufficiently flexible, shape-changing materials. With findings from NDS, perhaps researchers will eventually establish effective routes for making gels that act like cuttlefish. ■

Anna C. Balazs *is in the Chemical Engineering Department, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA.*
e-mail: balazs@pitt.edu

1. Gong, J. P. *Soft Matter* **6**, 2583–2590 (2010).
2. Zhou, H. *et al. Proc. Natl Acad. Sci. USA* **109**, 19119–19124 (2012).
3. Panyukov, S. & Rabin, Y. *Phys. Rep.* **269**, 1–131 (1996).
4. Yashin, V., Kuksenok, O., Dayal, P. & Balazs, A. C. *Rep. Progr. Phys.* **75**, 066601 (2012).
5. Cohen Stuart, M. A. *et al. Nature Mater.* **9**, 101–113 (2010).
6. Yoshida, R. *Sensors* **10**, 1810–1822 (2010).
7. He, X. *et al. Nature* **487**, 214–218 (2012).
8. Guillet, P. *et al. Soft Matter* **5**, 3409–3411 (2009).
9. Yashin, V. V., Kuksenok, O. & Balazs, A. C. *J. Phys. Chem. B* **114**, 6316–6322 (2010).
10. Ueno, T., Bundo, K., Akagi, Y., Sakai, T. & Yoshida, R. *Soft Matter* **6**, 6072–6074 (2010).

EARTH SCIENCE

How glaciers grow

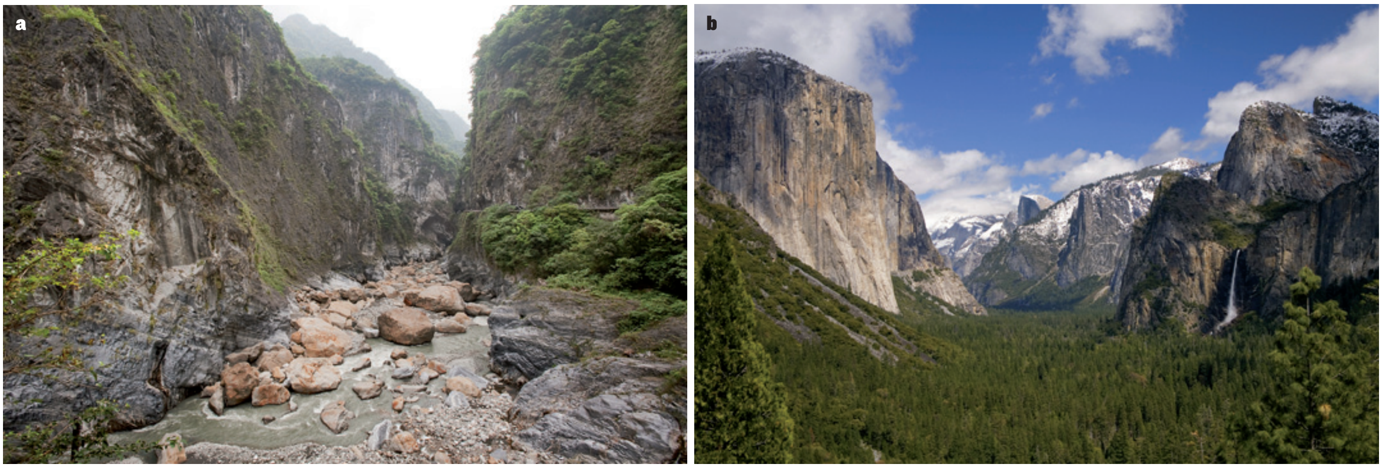
A state-of-the-art numerical model shows that the advance of glaciers in a cooling climate depends strongly on the pre-existing landscape, and that glacial erosion paves the way for greater glacial extent in the future. **See Letter p.206**

SIMON H. BROCKLEHURST

The growth of glaciers reflects the balance between the accumulation of snow and its loss through melting. Given the close association between altitude and temperature[1], the elevation and morphology of the valley floor on which a glacier forms are key determinants of a glacier's size and longevity. Glacial erosion has carved the spectacular alpine landscapes of many mountain ranges, and these are characterized by extensive glacial valley floors at elevations close to the long-term snowline[2]. These landscapes contrast strongly with their precursors, which are generally steeper, narrower valleys sculpted by rivers (Fig. 1). On page 206 of this issue, Pedersen and Egholm[3] use a state-of-the-art numerical model of glacier dynamics to quantify, for the first time, the stark contrast in glacier development between modern alpine mountain ranges and those that came before them. They compare the legacy of numerous glaciations during the Pleistocene epoch (between about 2.5 million and 10,000 years ago) with the landscapes that would have been present at the onset of these glaciations.

It has long been recognized that glaciers and their surroundings share an intimate, coupled relationship[4,5]. Shaded aspects and large areas at high elevation promote glacier growth. In turn, glacial erosion strongly modifies the underlying landscape, widening valley floors and making hill slopes steeper. However, given the efficiency of glaciers at reworking and removing the evidence left by previous glacial cycles, direct insight into the relationship between glaciers and topography during

JEFFREY YAP, JEFF RYAN STUDIO/GETTY



SEAN BAGSHAW/SPL

Figure 1 | The morphological contrast between fluvial and glacial landscapes. As documented by Pedersen and Egholm[3], the width and downstream gradient of a valley floor determine how glaciers would develop should the climate cool sufficiently. **a**, A fluvial landscape characterized by a narrow, steep valley: Taroko Valley, Hualien, Taiwan. Many of the world's mountain ranges would have been dominated by fluvial erosion before the onset of glaciations about 2.5 million years ago. The first glaciers to develop in such fluvial landscapes will show an almost linear relationship between temperature change and subsequent ice volume. **b**, A glacial landscape: Yosemite Valley, California. Ice volume will increase dramatically if the temperature cools sufficiently for the glacier to occupy the broad, shallow valley floor.

the initial phases of late Cenozoic glaciation (between about 2.5 million and 1 million years ago) has remained frustratingly elusive.

Numerical modelling offers the opportunity to explore glacier growth during initial glaciation. Pedersen and Egholm demonstrate an almost linear relationship between the degree of climate cooling and ice volume when glaciers develop in fluvial landscapes. However, for landscapes previously occupied and sculpted by glaciers, the equivalent relationship is highly nonlinear. Once the climate cools sufficiently, glaciers that descend onto the wide, shallow valley floor carved by preceding glacial occupations expand markedly.

Pedersen and Egholm also explore the full evolution of modern glacial landscapes by driving a simplified version of their numerical model — that is, without fluvial or hill-slope erosion, or active tectonics — with temperature fluctuations representing the past 2 million years. The mid-Pleistocene transition (MPT) about 950,000 years ago marked a shift from 40,000-year glacial cycles accompanied by symmetrical periods of cooling and warming to the protracted cooling and rapid warming of the more recent 100,000-year glacial cycles. The authors find much more extensive and erosive glaciers after the MPT than during the previous 1 million years. This result is consistent with geological evidence for accelerated glacial erosion after the MPT[6,7]. However, as Pedersen and Egholm show, this accelerated erosion is not simply a consequence of the changing climate; faster glacial erosion post-MPT was preconditioned by the landscape modifications made by the preceding, smaller-scale glaciations.

Pedersen and Egholm[3] have taken on a considerable challenge. Numerical modelling of landscape evolution resulting from glacial erosion is a complex, multi-faceted problem. As such, the first models of glacial landscape evolution have emerged only within the past 15 years[8]. Egholm and co-authors[3,9] have led the field and achieved compellingly realistic results by introducing ice-dynamics formulations that are more appropriate to valley glaciers — which descend the relatively steep downstream gradient of the valley floor and interact with the valley sides —than earlier simplifications borrowed from ice-sheet models.

Many considerations are involved in any attempt to numerically model glaciers and glacial erosion. The amount of snow that falls is strongly influenced by local climate, and is sensitive to changes in atmospheric moisture content, temperature and prevailing wind direction. Falling snow can be redistributed by wind[10] and avalanching before it reaches the glacier surface. Glacial-ice melt is influenced by temperature change, shading and rock fall from the valley sides. Glacial ice deforms under its own weight, and can also slide on its bed if liquid water is present. However, at this point, this generic model still only describes ice accumulating, deforming and melting. How would this moving ice (and water) erode the landscape?

As was recently noted[11], models of glacial landscape evolution, including Pedersen and Egholm's, still rely on simple, empirical relationships between glacier sliding velocity and glacial erosion rate, supported by modest field-data sets. There is little clear connection between these numerical relationships and the quarrying of large bedrock blocks from the glacier bed, which in most circumstances represents the primary process of glacial erosion. The liquid water in the subglacial hydrological network is also known to have a key role in glacial erosion, but determining the appropriate numerical formulation for the role of subglacial water is in its infancy[12]. Further progress will require numerical models that are robustly supported by a combination of careful observations of modern glaciers[13] and thermochronological evidence for how glacial landscapes have evolved over the longer timescales[7] that are explored numerically here. It will also require driving models with more-realistic climate representations[10,14] than those currently considered, and will probably depend on continuing advances in computer power. ■

Simon H. Brocklehurst *is in the School of Earth, Atmospheric and Environmental Sciences, The University of Manchester, Manchester M13 9PL, UK.*
e-mail: shb@manchester.ac.uk

1. Stone, P. H. & Carlson, J. H. *J. Atmos. Sci.* **36,** 415–423 (1979).
2. Brozović, N., Burbank, D. W. & Meigs, A. J. *Science* **276,** 571–574 (1997).
3. Pedersen, V. K. & Egholm, D. L. *Nature* **493,** 206–210 (2013).
4. Johnson, W. D. *J. Geol.* **12,** 569–578 (1904).
5. Gilbert, G. K. *J. Geol.* **12,** 579–588 (1904).
6. Haeuselmann, P., Granger, D. E., Jeannin, P.-Y. & Lauritzen, S.-E. *Geology* **35,** 143–146 (2007).
7. Valla, P. G., Shuster, D. L. & van der Beek, P. A. *Nature Geosci.* **4,** 688–692 (2011).
8. Braun, J., Zwartz, D. & Tomkin, J. H. *Ann. Glaciol.* **28,** 282–290 (1999).
9. Egholm, D. L., Knudsen, M. F., Clark, C. D. & Lesemann, J. E. *J. Geophys. Res.* **116,** F02012 (2011).
10. Anders, A. M., Roe, G. H., Montgomery, D. R. & Hallet, B. *Geology* **36,** 479–482 (2008).
11. Iverson, N. R. *Geology* **40,** 679–682 (2012).
12. Herman, F., Beaud, F., Champagnac, J.-D., Lemieux, J.-M. & Sternai, P. *Earth Planet. Sci. Lett.* **310,** 498–508 (2011).
13. Riihimaki, C. A., MacGregor, K. R., Anderson, R. S., Anderson, S. P. & Loso, M. G. *J. Geophys. Res.* **110,** F03003 (2005).
14. Rowan, A. V., Plummer, M. A., Brocklehurst, S. H., Jones, M. A. & Schultz, D. M. *Geology* http://dx.doi.org/10.1130/G33829.1 (2012).

LETTER

doi:10.1038/nature11630

Post-replicative repair involves separase-dependent removal of the kleisin subunit of cohesin

Alexandra McAleenan[1]*, Andres Clemente-Blanco[1]*, Violeta Cordon-Preciado[1], Nicholas Sen[1], Miguel Esteras[1], Adam Jarmuz[1] & Luis Aragón[1]

DNA double-strand break repair is critical for cell viability and involves highly coordinated pathways to restore DNA integrity at the lesion. An early event during homology-dependent repair is resection of the break to generate progressively longer 3′ single-strand tails that are used to identify suitable templates for repair[1]. Sister chromatids provide near-perfect sequence homology and are therefore the preferred templates during homologous recombination[2,3]. To provide a bias for the use of sisters as donors, cohesin—the complex that tethers sister chromatids together[4]—is recruited to the break[5,6] to enforce physical proximity. Here we show that DNA breaks promote dissociation of cohesin loaded during the previous S phase in budding yeast, and that damage-induced dissociation of cohesin requires separase, the protease that dissolves cohesion in anaphase[7]. Moreover, a separase-resistant allele of the gene coding for the α-kleisin subunit of cohesin, Mcd1 (also known as Scc1), reduces double-strand break resection and compromises the efficiency of repair even when loaded during DNA damage. We conclude that post-replicative DNA repair involves cohesin dissociation by separase to promote accessibility to repair factors during the coordinated cellular response to restore DNA integrity.

DNA double-strand breaks (DSBs), in which both strands of the DNA double helix are severed, are very hazardous as they can lead to genome rearrangements. In budding yeast, DSBs occurring post-replication are repaired by homologous recombination using sister chromatids as the template[2,3]. To facilitate homologous recombination between sister chromatids, the DNA-damage checkpoint promotes their physical proximity by allowing cohesion re-establishment[5,6,8,9]. In addition, the proximity between sister chromatids is also reinforced at the sites of damage through the formation of a *de novo* cohesin domain at breaks[5,6].

Cohesin complexes that are established in S phase and hold sister chromatids together have an unknown fate during the repair process. First, we wanted to know whether S-phase cohesin maintains cohesion after a DSB in G2 phase. An assay to measure damage-induced cohesion[9]—based on a green fluorescent protein (GFP) chromosome tag flanked by HO endonuclease sites that produce a fragment after DSB formation where cohesion can be analysed microscopically (Fig. 1a)—was modified to contain an *MCD1* gene regulated by the methionine promoter (*MET*). Passage through S phase in methionine-depleted media at 37 °C (to inactivate the *mcd1-1* allele) ensured that cohesion was established only by *MCD1* expressed from the *MET* promoter (Fig. 1a). Cells were blocked in metaphase before the expression of HO endonuclease (Fig. 1a). An increase in cohesion loss for the chromatin fragment was observed when *MCD1* transcription was repressed in metaphase after DSB induction (Fig. 1b). Therefore S-phase cohesin is unable to maintain cohesion of the fragment after a DSB. Next, we followed the dynamics of S-phase cohesin by western analysis. After establishment of cohesion, cells were blocked in metaphase and *MCD1* transcription was repressed with methionine. Upon DSB induction full-length Mcd1 disappeared (Fig. 1c); by contrast, a stable pool of Mcd1 was retained when the DSB was not induced. We conclude that the levels of Mcd1 loaded during S phase are markedly reduced as a consequence of DSB induction.

Next we asked whether constant transcription of *MCD1* is required post DSB formation to maintain cohesion. We substituted *MET* for the *GAL1-10* promoter in the strain with the GFP tags described above[9]. Expression of both *HO* and *MCD1*, promoted by galactose, was induced in metaphase cells (Supplementary Fig. 1). After DNA-damage-induced cohesion (Supplementary Fig. 1; cohesion establishment), cells were either maintained in galactose or transferred to glucose, which represses *GAL1-10*. Deletion of silent *HM* cassettes (donor sequences for the repair) in the strain ensured the persistence of the breaks in glucose. Cells transferred to glucose after cohesion establishment exhibited cohesion loss (Supplementary Fig. 1; glucose) compared with cells kept in galactose (Supplementary Fig. 1; galactose). Similar results were obtained with chromosome tags located on undamaged chromosomes (Supplementary Fig. 1). Moreover, Mcd1 binding dynamics around a *MAT* DSB after transient *MCD1* expression showed a reduction in binding upon transcriptional repression by glucose addition (Fig. 1d; glucose). Therefore an unused pool of Mcd1 is important in maintaining cohesion for extended periods in the presence of DNA damage. These results are consistent with a recent report demonstrating that DNA damage requires higher levels of cellular Mcd1 than cohesion establishment during S phase[10].

Next we sought to determine how Mcd1 dissociation is triggered. First, we considered Mcd1 de-acetylation, because Eco1-mediated acetylation promotes damage-induced cohesion[11]. We ruled out this possibility when we observed that an acetyl-mimicking allele coding for Mcd1 (Mcd1(K84/Q210)) lost cohesion at DSB-proximal regions after transcription repression (Supplementary Fig. 2). Next, we considered that separase might be required for the dissociation of cohesin after DNA damage, as in fission yeast, separase is required for DNA repair, and cleavage of Rad21, the Mcd1 homologue, has been reported in cells exposed to DNA-damage agents[12].

We used the assay described earlier (Fig. 1b) to test whether the reduction of S-phase cohesin levels requires separase. A tight conditional allele of yeast separase known as *esp1-C113* (ref. 13) was used to test this possibility. Unlike wild-type cells, *esp1-C113* cells maintained a stable pool of Mcd1 after DSB induction and *MCD1* transcriptional repression (Fig. 2a). Therefore, Mcd1 dissociation requires separase. The separase cleavage sites in Mcd1 have been mapped to arginine residues located at positions 180 and 268 (ref. 7). We mutated these residues to aspartic acid to generate an Mcd1 allele resistant to separase cleavage, *mcd1(RRDD)*[7]. Consistent with a role for separase in the dissociation of Mcd1, we found that transient expression of *mcd1(RRDD)* reduced both the loss of Mcd1 from DSB-proximal regions (Fig. 2b) and of damage-induced cohesion (Fig. 2c) observed for wild-type *MCD1* after transcription shutdown (Fig. 1c, d).

DNA damage causes hypermodification of Mcd1 (ref. 12), possibly by Chk1-dependent phosphorylation of Mcd1 serine residue 83 (ref. 14).

[1]Cell Cycle Group, MRC Clinical Sciences Centre, Imperial College, Du Cane Road, London W12 0NN, UK.
*These authors contributed equally to this work.

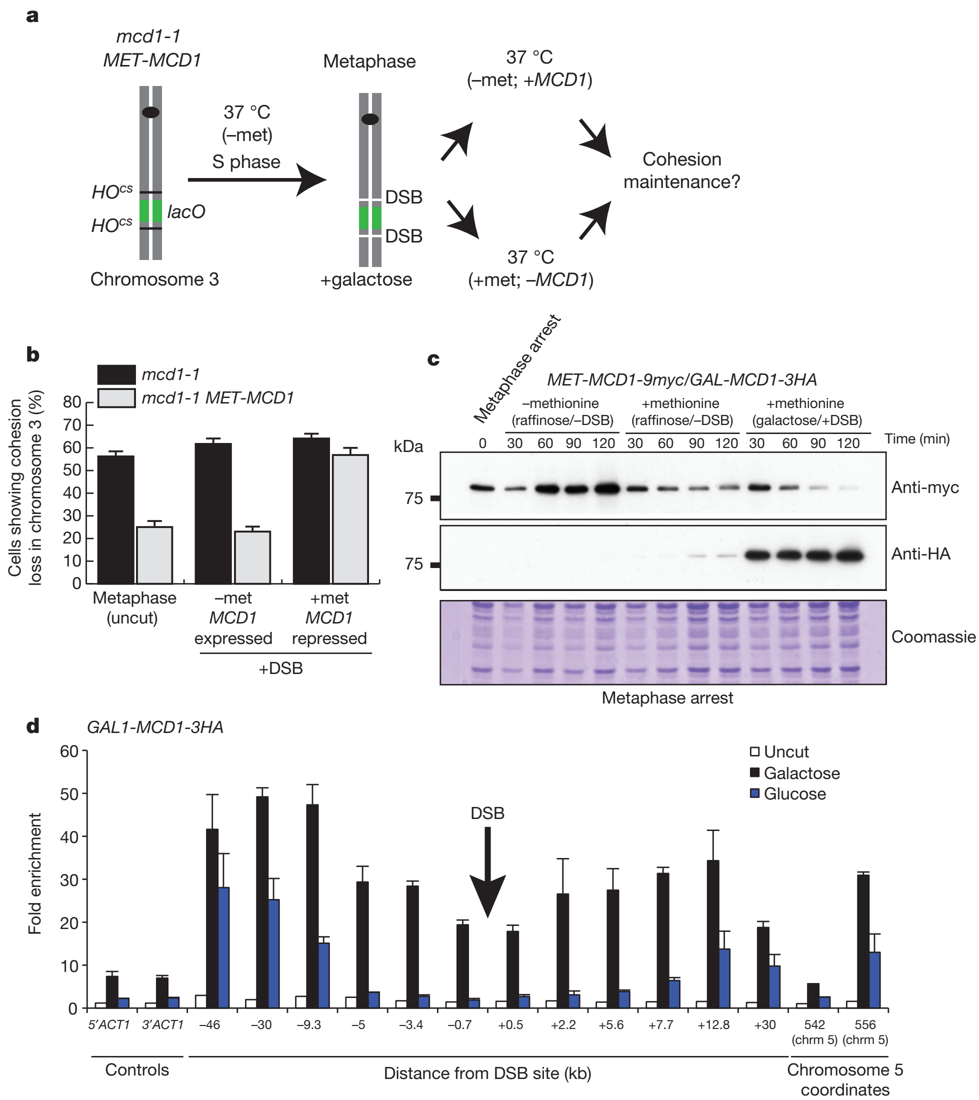


Figure 1 | S-phase-loaded cohesin is not sufficient to provide cohesion of a chromatin fragment in the presence of DNA damage. a, Diagrammatic representation of the modified assay[9] used to measure cohesivity of S-phase cohesin after DSB induction. Galactose was added to promote DSBs. *HO*^{cs}, *HO* cut site. **b**, Percentage of cells showing loss of cohesion is shown (mean ± s.e.m., $n = 3$), when incubated with or without methionine to repress further *MCD1* transcription. **c**, Western blot analysis of Mcd1 stability in cells containing the *GAL-HO* system, *MET-MCD1* attached to nine copies of the *myc* epitope tag (*MET-MCD1-9myc*) and *GAL-MCD1* attached to three copies of the haemagglutinin epitope tag (*GAL-MCD1-3HA*). **d**, Chromatin-immunoprecipitation (ChIP) analysis of Mcd1-3HA binding around HO-induced DSB at the *MAT* locus in cells with transient expression of Mcd1. Growth at 30 °C in raffinose media (uncut sample), followed by galactose addition for 3 h (expressing HO and Mcd1-3HA) and transfer to glucose media or maintenance on galactose. *ACT1* locus on chromosome 6 is shown for comparison (mean ± s.e.m., $n = 3$). The 556-kb position is a cohesin-association region.

We did not observe any change in mobility for S-phase-loaded Mcd1 when DSBs were induced in metaphase-arrested cells (Figs 1b and 2a). However, when Mcd1 was expressed simultaneously with HO endonuclease (generating DSBs), a slower migrating form of Mcd1 was observed (Supplementary Fig. 3). Phosphatase treatment demonstrated that this mobility change was due to protein phosphorylation (Supplementary Fig. 4). Interestingly, phosphorylated Mcd1 seemed more resistant to dissociation (Supplementary Fig. 3) and accumulated when separase-resistant Mcd1 was expressed simultaneously with the induction of a break (Supplementary Fig. 3). We also could detect what appeared to be a fragment of the carboxy-terminal region of Mcd1 (Supplementary Fig. 3), consistent with separase cleavage. On the basis of our findings we propose that in the presence of post-replicative DNA DSBs, S-phase and damage-induced Mcd1 can dissociate from DSB-proximal and cohesion-association regions in a separase-dependent manner. We speculate that this dissociation must be highly regulated and locally coordinated to prevent the full separation of the genome at any one time.

Nevertheless, even a localized dissociation of S-phase Mcd1 in cells experiencing post-replicative DNA damage is surprising, as cohesion is thought to be required for accurate repair. It is possible however, that local cohesion removal is necessary for repair. DSB repair by homologous recombination initiates with 5′–3′ resection of the break to yield 3′ single-stranded DNA tails[1]. In budding yeast, resection is initiated by the MRX (Mre11–Rad50–Xrs2) complex[15,16], and more extensive resection is accomplished by the combined activities of the Sgs1–Top3–Rmi1 complex and Dna2 (ref. 16) and/or the exonuclease Exo1 (ref. 17). The MRX subunit Mre11 localized to DSBs when cohesin dissociation was blocked in cells expressing the Mcd1 separase-resistant allele (Supplementary Fig. 5); however, Sgs1 recruitment was reduced in these cells (Fig. 3a), raising the possibility that break resection might be slower or inefficient in the absence of Mcd1 dissociation. To evaluate this, we compared break-resection dynamics in cells expressing wild-type or separase-resistant Mcd1 (Fig. 3b). After synchronous HO-induced cleavage, the rate of resection was evaluated

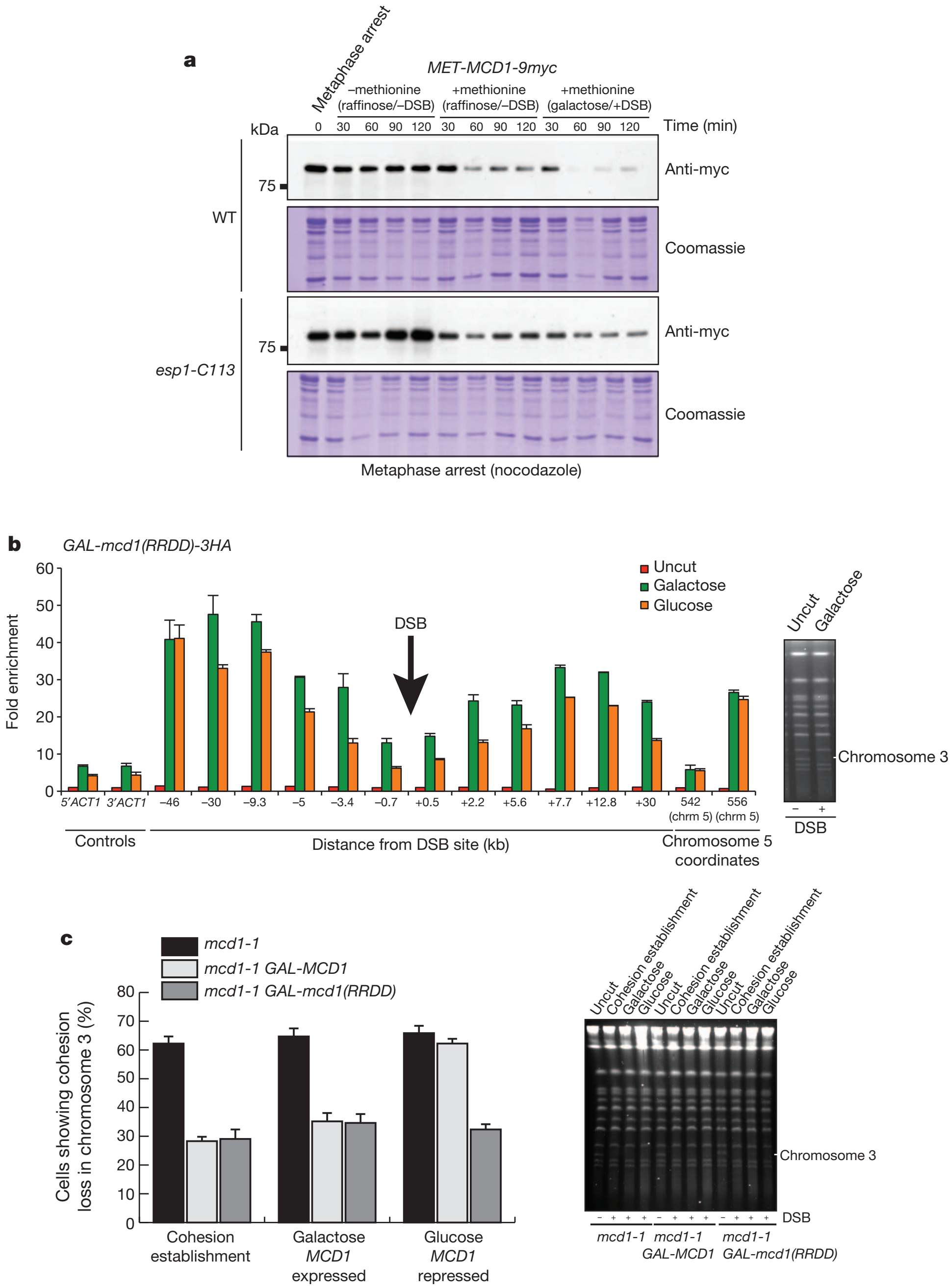


Figure 2 | Mcd1 dissociation in the presence of DNA damage requires separase. **a**, Western blot analysis of Mcd1 stability in wild-type and *esp1-C113* cells carrying the *GAL-HO* system and *MET-MCD1-9myc*. Coomassie stain shows equal loading on the blots. **b**, ChIP analysis of Mcd1(RRDD)-3HA binding around the HO-induced DSB at the *MAT* locus in cells during the growth conditions described in Fig. 1d, before DSB (uncut), after the addition of galactose for 5 h (galactose) and after the addition of galactose for 3 h followed by transfer to glucose for 2 h (glucose). Binding at the *ACT1* locus on chromosome 6 is shown for comparison (mean ± s.e.m., $n = 3$). Pulse-field gel electrophoresis to confirm DSB occurrence is shown. **c**, Percentage of *mcd1-1*, *mcd1-1/GAL-MCD1* and *mcd1-1/GAL-mcd1(RRDD)* cells showing loss of cohesion (as described in Fig. 1d,) after 2 h of galactose addition (cohesion establishment), after 3.5 h in galactose (galactose; *MCD1* expressed) and after 2 h in galactose followed by 1.5 h in glucose (glucose; *MCD1* repressed) (mean ± s.e.m, $n = 4$) (left). Pulse-field gel electrophoresis to confirm DSB occurrence is shown (right).

using probes at different distances from the HO break. As the 5′ strand is degraded, restriction enzymes are unable to cleave single-stranded DNA, and the intensity of the bands corresponding to the DNA fragments by Southern blot hybridization becomes diminished. Resection in cells expressing separase-resistant Mcd1 was attenuated (Fig. 3b and Supplementary Fig. 6). A similar result was found when resection intermediates were analysed as transient smears of higher molecular weight (Supplementary Fig. 7). These results demonstrate that, in the absence of Mcd1 dissociation, cohesin complexes can interfere with resection of 5′ termini.

Next, we tested whether blocking Mcd1 dissociation impairs DNA-repair events that require extended end resection. To this aim, we used a yeast strain engineered to contain a 117-base-pair HO-endonuclease-recognition sequence within the *LEU2* gene on chromosome 3, and carrying homology with only the right end of the break (the 'U2' end) 25 kilobases (kb) away[18,19] (Supplementary Fig. 8). Repair largely occurs through single-strand annealing[19], but also can occur by Rad51-dependent intra-chromosomal break-induced replication (Supplementary Fig. 8). Cells expressing separase-resistant Mcd1 exhibited a reduction in repair product (Fig. 4a), demonstrating

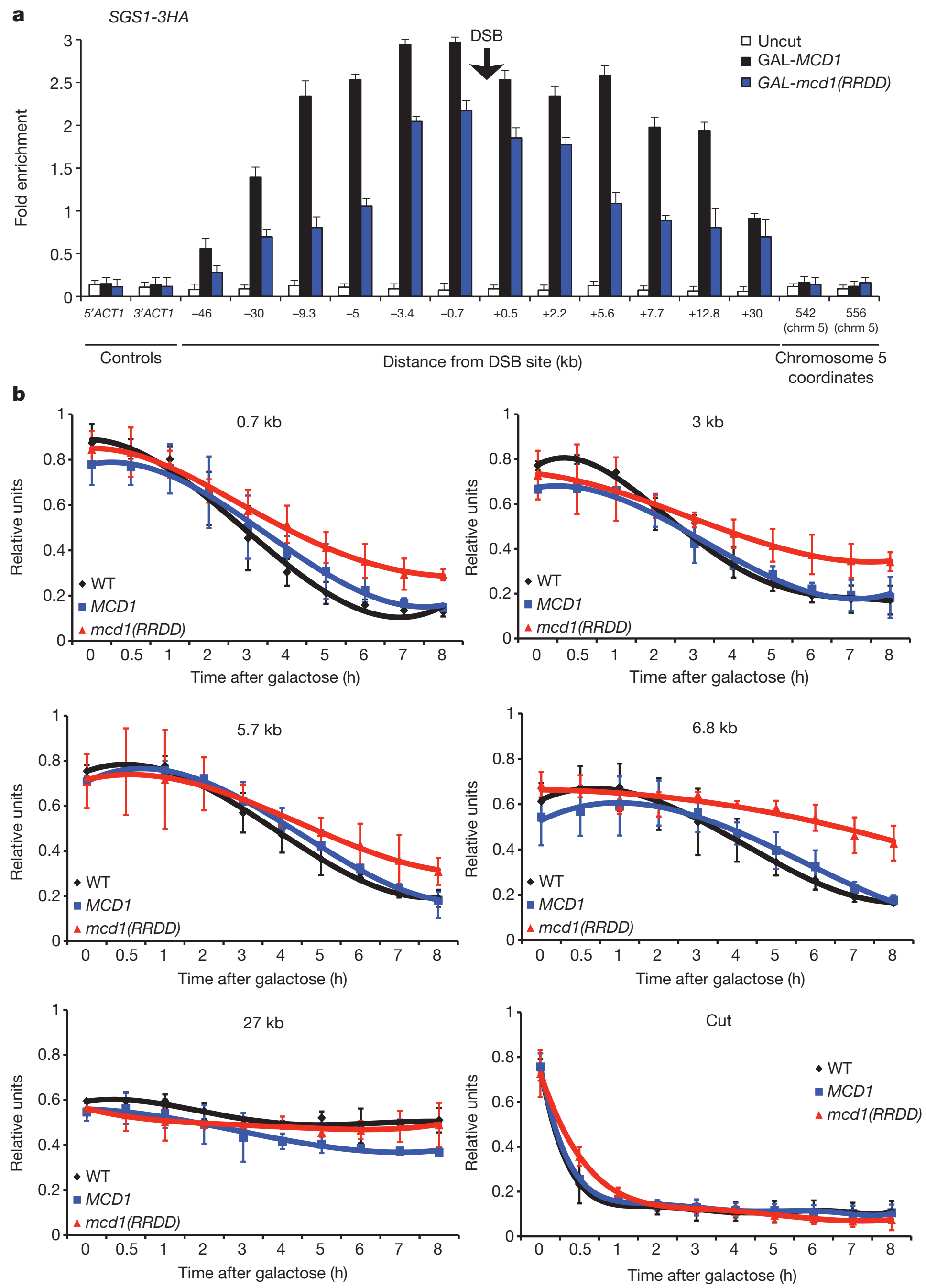


Figure 3 | Mcd1 dissociation facilitates 5′–3′ end resection. a, ChIP analysis of Sgs1 binding around the HO-induced DSB at the *MAT* locus in cells expressing wild-type Mcd1 or Mcd1(RRDD). Cells carrying *SGS1-3HA* under the endogenous promoter and the *HO* endonuclease system together with either *GAL-MCD1* or *GAL-mcd1(RRDD)* were grown at 30 °C in raffinose media (uncut), and expression of *HO* and either *MCD1* or *mcd1(RRDD)* was induced for 3 h. Binding at the *ACT1* locus on chromosome 6 is shown for comparison (mean ± s.e.m., $n = 3$). **b**, Plots show quantification of band intensities corresponding to each probe used over time on Southern blot analysis measuring strand resection in wild-type cells expressing an empty vector (*MCD1*; WT), *GAL-MCD1* or *GAL-mcd1(RRDD)* (mean ± s.e.m., $n = 3$).

that Mcd1 dissociation is important for efficient repair of the break in this assay. Expression of separase-resistant Mcd1 in a *rad51Δ* derivative strain, in which break-induced replication is inactive, also exhibited reduced repair kinetics (Supplementary Fig. 8). Moreover, analysis of *MAT*-switching kinetics—which does not involve extensive resection—in cells expressing separase-resistant Mcd1 showed that although switching was efficiently achieved, the formation of recombinant product was delayed compared with wild-type Mcd1 (Supplementary Fig. 9). These data demonstrate that local cohesin dissociation facilitates the timely repair of DNA DSBs.

On the basis of our results we propose that upon DSB formation post-replication, cohesin complexes loaded during S phase, and to some extent those loaded *de novo* at DSB-proximal regions, dissociate from chromosomes (Supplementary Fig. 10). Importantly, cohesin dissociation involves separase, hence preventing the use of dissociated Mcd1 in a new round of cohesion generation. The functional role of cohesin dissociation is to allow local accessibility for repair activities, such as resection, at sites of damage (Supplementary Fig. 10). Separase activity must be tightly regulated during DNA repair, perhaps with targeted or attenuated proteolytic activity (compared to anaphase), to ensure context-dependent cleavage of Mcd1. The detailed mechanisms

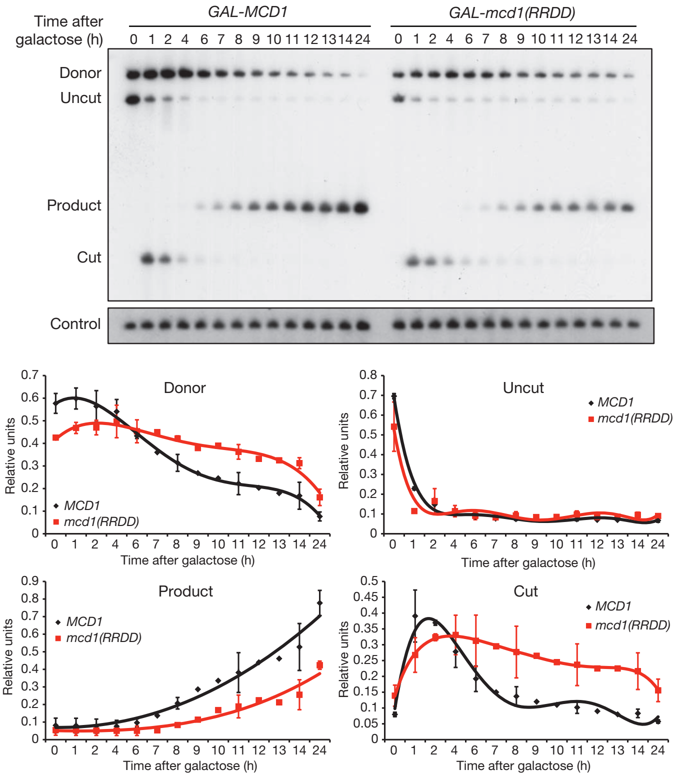


Figure 4 | Mcd1 dissociation facilitates repair of DNA DSB. a, Top, Southern blot showing kinetics of repair in strains carrying the donor U2 sequence 25 kb away from the *HO* cut site and expressing either a wild-type copy of Mcd1 under the *GAL1-10* promoter (*GAL-MCD1*) or a non-cleavable mutant of Mcd1 also under *GAL1-10* promoter control (*GAL-mcd1(RRDD)*). Cells were first grown at 30 °C in raffinose media (uncut) before expression of *HO* and either *MCD1* or *mcd1(RRDD)* by addition of galactose was induced. Samples were collected at the indicated time points and DNA was digested with KpnI and probed with a U2-specific probe. The bands representing donor, uncut, product and cut species are indicated. Bottom, plots show quantification of band intensities corresponding to each band (donor, uncut, product and cut) over time (mean ± s.e.m., $n = 3$).

of this regulation are a fascinating question for the future. Interestingly, if cohesin dissociation is conserved in mammalian models, therapeutic approaches that rely on the generation of DNA lesions could be improved by simultaneously inhibiting the cohesin cycle.

METHODS SUMMARY

Induction of DSBs and shut-off assays. DSBs were induced in strains containing stably integrated *GAL1-10::HO* and *MCD1-3*(or *-6*)*::HA* alleles, either on centromeric plasmids (carrying the *LEU2* marker) or stably integrated. Cells were grown in YEP media with 2% lactic acid or 2% raffinose. Galactose (Sigma, 20% w/v stock) was added to the culture at a final concentration of 2% w/v. For ChIP experiments cells were incubated in galactose, to induce alleles coding for HO and Mcd1, for 4 h at 30 °C, and the culture was then split in two. One half was maintained in 2% galactose, whereas the remainder of the culture was incubated in 2% glucose. Samples were collected at different time points (as indicated in figure legends) and processed for ChIP. For experiments to analyse damage-induced cohesion, cells were arrested in metaphase with 15 μg ml^{-1} nocodazole (1:100 dilution from 1.5 mg ml^{-1} in dimethylsulphoxide (DMSO)) before addition of galactose to induce HO and Mcd1 expression, and were incubated for 1 h at 30 °C, then transferred to 37.5 °C to inactivate *mcd1-1*. After incubation a second dose of nocodazole was added (15 mg ml^{-1}) to prevent anaphase entry. One half of the culture was then maintained in 2% galactose, whereas the remainder of the culture was incubated in 2% glucose. Samples were collected at different time points (as indicated in figure legends) and processed for microscopy. For extended arrests in metaphase a second dose of nocodazole was added after 3 h to prevent cells entering anaphase.

DSB resection and SSA assays. The restriction enzyme used for 5′–3′ resection analysis was StyI (New England Biolabs). The digested fragments were resolved on agarose gels and transferred to GeneScreen Plus hybridization membrane either by vacuum or capillary wet transfer. The DNA probes used for Southern blotting were prepared by Amersham DNA labelling beads (−dCTP) and purified by Amersham G-50 Micro Columns. For single-strand annealing experiments the restriction enzyme used was KpnI.

Full Methods and any associated references are available in the online version of the paper.

Received 9 December 2011; accepted 28 September 2012.
Published online 25 November 2012.

1. Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev.* **63,** 349–404 (1999).
2. Kadyk, L. C. & Hartwell, L. H. Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae. Genetics* **132,** 387–402 (1992).
3. Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J.* **19,** 3398–3407 (2000).
4. Nasmyth, K. Cohesin: a catenase with separate entry and exit gates? *Nature Cell Biol.* **13,** 1170–1177 (2011).
5. Unal, E. *et al.* DNA damage response pathway uses histone modification to assemble a double-strand break-specific cohesin domain. *Mol. Cell* **16,** 991–1002 (2004).
6. Strom, L., Lindroos, H. B., Shirahige, K. & Sjogren, C. Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. *Mol. Cell* **16,** 1003–1015 (2004).
7. Uhlmann, F., Lottspeich, F. & Nasmyth, K. Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature* **400,** 37–42 (1999).
8. Strom, L. *et al.* Postreplicative formation of cohesion is required for repair and induced by a single DNA break. *Science* **317,** 242–245 (2007).
9. Unal, E., Heidinger-Pauli, J. M. & Koshland, D. DNA double-strand breaks trigger genome-wide sister-chromatid cohesion through Eco1 (Ctf7). *Science* **317,** 245–248 (2007).
10. Heidinger-Pauli, J. M., Mert, O., Davenport, C., Guacci, V. & Koshland, D. Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. *Curr. Biol.* **20,** 957–963 (2010).
11. Heidinger-Pauli, J. M., Unal, E. & Koshland, D. Distinct targets of the Eco1 acetyltransferase modulate cohesion in S phase and in response to DNA damage. *Mol. Cell* **34,** 311–321 (2009).
12. Nagao, K., Adachi, Y. & Yanagida, M. Separase-mediated cleavage of cohesin at interphase is required for DNA repair. *Nature* **430,** 1044–1048 (2004).
13. Baskerville, C., Segal, M. & Reed, S. I. The protease activity of yeast separase (esp1) is required for anaphase spindle elongation independently of its role in cleavage of cohesin. *Genetics* **178,** 2361–2372 (2008).
14. Heidinger-Pauli, J. M., Unal, E., Guacci, V. & Koshland, D. The kleisin subunit of cohesin dictates damage-induced cohesion. *Mol. Cell* **31,** 47–56 (2008).
15. Zhu, Z., Chung, W. H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* **134,** 981–994 (2008).
16. Mimitou, E. P. & Symington, L. S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455,** 770–774 (2008).
17. Shim, E. Y. *et al.* Saccharomyces cerevisiae Mre11/Rad50/Xrs2 and Ku proteins regulate association of Exo1 and Dna2 with DNA breaks. *EMBO J.* **29,** 3370–3380 (2010).
18. Vaze, M. B. *et al.* Recovery from checkpoint-mediated arrest after repair of a double-strand break requires Srs2 helicase. *Mol. Cell* **10,** 373–385 (2002).
19. Jain, S. *et al.* A recombination execution checkpoint regulates the choice of homologous recombination pathway during DNA double-strand break repair. *Genes Dev.* **23,** 291–303 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank D. Koshland, E. Unal, J. Haber, G. Ira and K. Nasmyth for providing valuable strains. We thank members of the Aragon laboratory, P. Ullal, S. Farmer, K. Pegram, J. Torres-Rosell and M. Merkenschlager for discussions and reading of the manuscript. This work was funded by the Intramural Research Program of the Medical Research Council UK.

Author Contributions A.M. made the original observation that transcriptional repression of *MCD1* in the presence of DNA damage causes defects in sister chromatid cohesion. A.M. and A.C.-B. performed all ChIP analysis. A.M., A.C.-B., N.S. and V.C.-P. performed microscopic analysis of cohesion. A.C.-B. and M.E. performed physical assays to measure resection and repair. A.M. and A.C.-B. performed biochemical assays to measure Mcd1 stability. A.J. provided technical support. L.A. wrote the paper, and all authors analysed the data, discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.A. (luis.aragon@csc.mrc.ac.uk).

METHODS

Yeast strains and DNA constructs. Strains used in this study are isogenic to BY4741 (*MAT**a***, *his3Δ1*, *leu2Δ0*, *met15Δ0*, *ura3Δ0*) or W303 (*MAT**a***, *can1-100*, *leu2-3*, *his3-11*, *trp1-1*, *ura3-1*, *ade2-1*), except for the strains with the HO system (obtained from J. Haber). Strains used are listed in Supplementary Table 1. All DNA constructions were done using Gene-synthesis service from GeneCust.

Yeast media and cell cycle synchronization. For G1 arrests, α-factor was added to exponentially growing *MAT**a*** cultures ($D_{600\,nm} \approx 0.5$) to a final concentration of 1.5×10^{-8} M for *bar1Δ* strains and 3×10^{-6} M for *BAR1* strains. Cultures were monitored by microscopy until ≥90% of cells were arrested in G1 (typically 2–3 h). To inactivate temperature-sensitive alleles in G1, cells were shifted to 37 °C for 30 min before release. To release cells, the culture was spun (4,000 r.p.m., 2 min) and the pellet washed three times in a large volume of pre-warmed medium containing 0.1 mg ml^{-1} pronase. The pellet was then re-suspended in fresh medium and samples taken at the indicated times. Nocodazole was used to arrest cells in G2/M (1.5 mg ml^{-1} stocks in dimethylsulphoxide (DMSO)). To repress transcription from the methionine promoter we added 5 mM methionine to the cultures.

Induction of DSBs and shut-off assays. DSBs were induced in strains containing stably integrated *GAL1-10::HO* and *MCD1-3*(or *-6*)*::HA* alleles, either on centromeric plasmids (carrying the *LEU2* marker) or stably integrated. Cells were grown in YEP media with 2% lactic acid or 2% raffinose. Galactose (Sigma, 20% w/v stock) was added to the culture at a final concentration of 2% w/v. For ChIP experiments cells were incubated in galactose to induce alleles coding for HO and Mcd1 for 4 h at 30 °C, the culture was then split in two. One half was maintained in 2% galactose whereas the remainder of the culture was incubated in 2% glucose. Samples were collected at different time points (as indicated in figure legends) and processed for ChIP. For experiments to analyse damage-induced cohesion, cells were arrested in metaphase with 15 μg ml^{-1} nocodazole (1:100 dilution from 1.5 mg ml^{-1} in DMSO) before addition of galactose to induce HO and Mcd1 expression, and were incubated for 1 h at 30 °C, then transferred to 37.5 °C to inactivate *mcd1-1*. After incubation a second dose of nocodazole was added (15 μg ml^{-1}) to prevent anaphase entry. One half of the culture was then maintained in 2% galactose whereas the remainder of the culture was incubated in 2% glucose. Samples were collected at different time points (as indicated in figure legends) and processed for microscopy. For extended arrests in metaphase a second dose of nocodazole was added after 3 h to prevent cells entering anaphase.

Microscopy. For microscopy, series of *Z*-focal plane images were collected on a Leica IRB microscope using a Hamamatsu 742-95 digital camera and OpenLab software (Improvision). A tunable light source (Polychrome IV (Photonics)) with a xenon lamp or an ultraviolet mercury lamp (Leica) was used. Images in different *z* axis planes were flattened into a two-dimensional projection and processed in OpenLab. To visualize the nuclei of intact cells, cells were re-suspended in a final concentration of 1% Triton X-100 and 25 ng ml^{-1} DAPI/Antifade. For each experimental condition, at least 200 cells were scored for GFP spots.

ChIP. For ChIP analysis, 50 ml of culture (at $D_{600\,nm} \approx 0.5$) was fixed with formaldehyde (final concentration, 1.42%) for 15 min at room temperature (20 °C). The formaldehyde was quenched with glycine (final concentration, 125 mM) for 5 min at room temperature and the cells collected by centrifugation (4,000 r.p.m., 2 min). The pellet was washed with PBS buffer, transferred to a screw-cap tube and snap-frozen in liquid nitrogen. Pellets were stored at −80 °C. Pellets were re-suspended in 100 μl IP buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 5 mM EDTA, NP-40 (0.5% v/v), Triton X-100 (1.0% v/v)) containing phenylmethylsulphonyl fluoride (PMSF) (final concentration, 1 mM) and complete protease inhibitor cocktail (Roche), and 500 μl glass beads were added. Cells were lysed by two 20-s cycles on power setting 6.5 in a FastPrep FP120 (BIO 101) machine, with 5 min on ice in between cycles. 300 μl IP buffer containing PMSF and protease inhibitors was then added. Tubes were pierced with a hot needle and placed on top of fresh eppendorfs and spun (1,000 r.p.m., 2 min) to collect lysate minus glass beads. Samples were pelleted (13,000 r.p.m., 1 min, 4 °C). The nuclear pellet was re-suspended thoroughly in 1 ml IP buffer containing PMSF and protease inhibitors. The chromatin was sonicated for 1 h (15 s on, 15 s off at high power at 4 °C) (Diagenode Bioruptor). At this point, sonicated chromatin could be stored at −80 °C. 120 μl (30% of sample) was taken for input and 400 μl for IP. After clarification (13,000 r.p.m., 1 min) input DNA was precipitated from the supernatant by the addition of KaOAc to a final concentration of 0.3 M and 2.75 volumes of ethanol. The mixture was incubated at −20 °C for 2 h or overnight, and then spun (13,000 r.p.m., 5 min) and the supernatant discarded. 100 μl 10% (w/v) Chelex 100 suspension was added to the dried pellet and the sample boiled for 20 min. The supernatant was cleaned using the PCR purification kit (Qiagen) according to the manufacturer's instructions, and the DNA eluted in 250 μl water. DNA was stored at −20 °C. 2 μg anti-HA antibody (Roche) was added to 400 μl chromatin for the IP, and the samples were incubated in an ultrasonic water bath for 30 min at 4 °C. After clarification (13, 000 r.p.m., 2 min, 4 °C) the supernatant was added to 100 μl of a 50:50 slurry of Protein A and Protein G beads (Roche), which had been pre-equilibrated in IP buffer. The sample and beads were incubated for 2 h or overnight on a rotating platform at 4 °C. The beads were then washed three times in IP buffer, and then 250 μl 10% (w/v) Chelex 100 suspension added and the sample boiled for 20 min. After spinning down (13,000 r.p.m., 1 min) the supernatant was transferred into a fresh tube and stored at −20 °C.

Real-time PCR. PCR reactions were performed using the Sensimix NoRef Kit (Quantance). Reactions were carried out according to the manufacturer's instructions in a total volume of 20 μl containing 3 μl input or IP DNA and oligonucleotide primer pairs (final concentration, 1.5 μM). Amplification was performed in a DNA Engine Opticon 2 thermal cycler and analysed using Opticon software (MJ Research) or in a Bio-Rad C1000 Thermal Cycler in conjunction with the Bio-Rad CFX96 Real-time-system and analysed using CFX manager (Bio-Rad). The primers used for the analysis of an HO break at the *MAT* locus are shown in Supplementary Table 2.

Sample preparation and western blot analysis. Samples were prepared by trichloroacetic acid (TCA) extraction. Extracts were prepared as follows: cells were collected by centrifugation (4,000 r.p.m., 2 min) and washed with 20% TCA. The TCA was aspirated and the pellets frozen at −80 °C. All of the following purification steps were performed on ice with pre-chilled solutions. Cells were re-suspended in 250 μl 20% TCA, glass beads were added and the cells broken by one 40-s cycle at power setting 5.5 in a FastPrep FP120 (BIO 101) machine. Tubes were pierced with a hot needle and placed onto of fresh eppendorfs and spun (1,000 r.p.m., 2 min) to collect lysate minus glass beads. The glass beads were washed with 1 ml 5% TCA and this was added to the lysate, and mixed by pipetting. The precipitated proteins were collected by centrifugation (14,000 r.p.m., 10 min, 4 °C) and then pellets were washed with 750 μl 100% ethanol. Proteins were solubilized in 50 μl 1M Tris, pH 8, and 100 μl (×2) SDS–PAGE loading buffer (60 mM Tris, pH 6.8, 2% SDS, 10% glycerol, 0.2% bromophenol blue) and boiled for 5 min at 95 °C. Insoluble material was removed by centrifugation (14,000 r.p.m., 5 min, room temperature) and the supernatant either stored at −20 °C or loaded immediately onto an SDS–PAGE mini-gel. Samples were either run on an 8% acrylamide gel in Tris-Glycine SDS running buffer using the Bio-Rad Mini-PROTEAN 3 system or on Pre-Cast 4–12% Bis-Tris gels in NuPAGE MOPS (all Invitrogen). SDS–PAGE gels were transferred to polyvinylidene fluoride transfer membrane (Hybond-P, Amersham Biosciences) in either the Bio-Rad Mini Trans-Blot Electrophoretic Transfer Cell or by using the XCell SureLock Mini Cell transfer module (Invitrogen). The Bio-Rad system was used in conjunction with Tris-Glycine blotting buffer (National Diagnostics) containing 20% methanol, and run for 1 h at 200 V or overnight at 30 V. The NuPAGE system was used with NuPAGE transfer buffer containing 20% methanol and run for 1 h at 30 V. Membranes were blocked in 5% skimmed-milk powder in PBS with 0.1% Tween 20 (PBS-T) for 1 h or overnight at 4 °C, then incubated with the anti-HA IgG1 antibody 12CA5 (Roche) at a 1/5,000 dilution in blocking solution for between 1 h at room temperature to overnight at 4 °C. After several washes in PBS-T, membranes were incubated with the sheep anti-mouse IgG horseradish peroxidase-linked antibody (GE Healthcare) at a 1/10,000 dilution in blocking solution. After several further washes in PBS-T, membranes were incubated with the ECL Plus Western Blotting Detection system (GE Healthcare), followed by exposure to ECL Hyperfilm (GE Healthcare) to detect the secondary antibody.

Pulse-field gel electrophoresis. Pulse-field gel electrophoresis was performed by running genomic DNA embedded in plugs of 1% agarose, at 200 V, for 40 h (initial time, 10 s; final time, 35 s). DNA plugs were made using a Bio-Rad CHEF yeast genomic DNA plug kit, according to the manufacturer's protocol.

DSB resection and single-strand annealing assays. Genomic DNA was prepared as described. The restriction enzyme used for 5′–3′ resection analysis was StyI (New England Biolabs). The digested fragments were resolved on agarose gels and transferred to GeneScreen Plus hybridization membrane either by vacuum or by capillary wet transfer. The DNA probes used for Southern blotting were prepared by Amersham DNA-labelling beads (−dCTP) and purified by Amersham G-50 Micro Columns. For single-strand annealing experiments the restriction enzyme used was KpnI. The experiment was performed as described in refs 18 and 19.

doi:10.1038/nature11661

Control of somatic tissue differentiation by the long non-coding RNA TINCR

Markus Kretz[1], Zurab Siprashvili[1]*, Ci Chu[1]*, Dan E. Webster[1]*, Ashley Zehnder[1], Kun Qu[1], Carolyn S. Lee[1], Ross J. Flockhart[1], Abigail F. Groff[1], Jennifer Chow[1], Danielle Johnston[1], Grace E. Kim[1], Robert C. Spitale[1], Ryan A. Flynn[1], Grace X. Y. Zheng[1], Subhadra Aiyer[2], Arjun Raj[2], John L. Rinn[3], Howard Y. Chang[1,4] & Paul A. Khavari[1,5]

Several of the thousands of human long non-coding RNAs (lncRNAs) have been functionally characterized[1–4]; however, potential roles for lncRNAs in somatic tissue differentiation remain poorly understood. Here we show that a 3.7-kilobase lncRNA, terminal differentiation-induced ncRNA (TINCR), controls human epidermal differentiation by a post-transcriptional mechanism. TINCR is required for high messenger RNA abundance of key differentiation genes, many of which are mutated in human skin diseases, including *FLG, LOR, ALOXE3, ALOX12B, ABCA12, CASP14* and *ELOVL3.* TINCR-deficient epidermis lacked terminal differentiation ultrastructure, including keratohyalin granules and intact lamellar bodies. Genome-scale RNA interactome analysis revealed that TINCR interacts with a range of differentiation mRNAs. TINCR–mRNA interaction occurs through a 25-nucleotide 'TINCR box' motif that is strongly enriched in interacting mRNAs and required for TINCR binding. A high-throughput screen to analyse TINCR binding capacity to approximately 9,400 human recombinant proteins revealed direct binding of *TINCR* RNA to the staufen1 (STAU1) protein. STAU1-deficient tissue recapitulated the impaired differentiation seen with TINCR depletion. Loss of *UPF1* and *UPF2*, both of which are required for STAU1-mediated RNA decay, however, did not have differentiation effects. Instead, the TINCR–STAU1 complex seems to mediate stabilization of differentiation mRNAs, such as *KRT80.* These data identify TINCR as a key lncRNA required for somatic tissue differentiation, which occurs through lncRNA binding to differentiation mRNAs to ensure their expression.

LncRNAs regulate a variety of processes[1–3,5], yet their effects on homeostasis in somatic tissues such as epidermis are not fully defined. Transcriptome sequencing of progenitor and differentiating human keratinocytes was undertaken using the Illumina paired-end HiSeq platform with a read length of 101 nucleotides at about 110 million mapped reads per sample, with consistency verified by quantitative reverse transciptase PCR (qRT–PCR) (Supplementary Fig. 1a–c). TINCR, an uncharacterized lncRNA[6], was among the most highly induced lncRNAs of the 258 annotated non-coding RNAs changing during differentiation (Fig. 1a and Supplementary Table 1). The *TINCR* gene resides on chromosome 19 in humans between the *SAFB2* and *ZNRF4* genes (Fig. 1b), a locus conserved with a syntenic non-coding region on mouse chromosome 17qD. *TINCR* produces a 3.7-kilobase (kb) transcript induced >150-fold during epidermal differentiation (Fig. 1c–e). *TINCR* is downregulated in human squamous

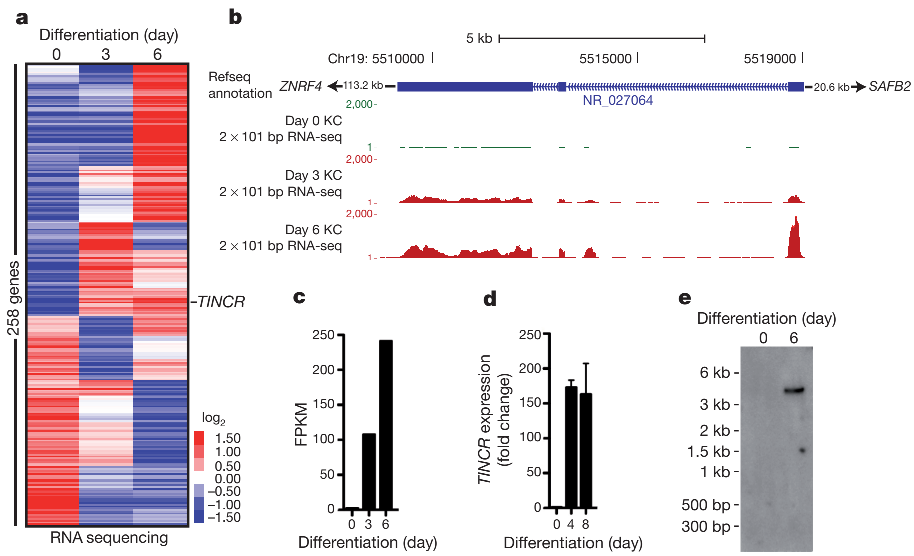


Figure 1 | TINCR is induced during epidermal differentiation. **a**, Mean-centred, hierarchical clustering of 258 annotated non-coding RNAs altered (>twofold change) in undifferentiated cells (day 0) and during days of calcium-induced differentiation *in vitro*. **b**, Schematic of *TINCR* genomic locus on chromosome 19. Day 0, 3 and 6 of keratinocyte (KC) differentiation; blue rectangles represent exons. **c**, Relative TINCR abundance in fragments per kilobase of exon model per million mapped fragments (FPKM). **d**, *TINCR* qRT–PCR. Error bars are s.d., $n = 4$. **e**, Northern blot analysis, with TINCR the single band seen in differentiation; bp, base pairs.

[1]The Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA. [2]Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [4]Howard Hughes Medical Institute, Stanford, California 94305, USA. [5]Veterans Affairs Palo Alto Healthcare System, Palo Alto, California 94304, USA.
*These authors contributed equally to this work.

cell carcinoma specimens, consistent with decreased differentiation seen in squamous cell carcinomas (Supplementary Fig. 1d). Single-molecule RNA fluorescence *in situ* hybridization (FISH) identified 80.6% of TINCR molecules newly acquired during differentiation within the cytoplasm (Supplementary Fig. 1e, f). FISH in human epidermis showed enrichment of TINCR in differentiated layers (Supplementary Fig. 1g). TINCR is therefore a differentiation-induced, predominantly cytoplasmic lncRNA.

TINCR function was assessed by RNA interference in organotypic human epidermal tissue, a setting that recapitulates the structure and gene expression of human epidermis[7,8]. Although TINCR-deficient epidermis stratified normally, the expression of key differentiation genes mutated in human diseases of abnormal epidermal function[9–11] was markedly reduced at the protein (Fig. 2a) and mRNA (Fig. 2b) levels. TINCR is thus required for normal induction of key protein mediators of epidermal differentiation.

Transcript profiling of TINCR-depleted epidermis demonstrated that TINCR loss disrupted the expression of 394 genes (Supplementary Fig. 2a and Supplementary Table 2). TINCR-regulated genes were enriched for differentiation-associated epidermal barrier formation-related Gene Ontology (GO) terms (Fig. 2c). Barrier formation requires genes encoding the protein structure of the terminally differentiated stratum corneum, such as loricrin and filaggrin, as well as those synthesizing specific water-impermeable lipids[12]. GO terms related to the latter were enriched in genes altered by TINCR loss, as were the mRNA levels of genes in this subset that are genetically nonredundant for epidermal barrier formation[13–15] (Fig. 2d). Furthermore, caspase 14, implicated in proteolysis needed for epidermal barrier function[16], was diminished by 83.7% with TINCR loss. Protein and lipid barrier ultrastructures involved in barrier formation were abnormal in the outer layers of TINCR-deficient epidermis, including protein-rich keratohyalin granules (Fig. 2e) and the lipid-rich lamellar bodies (Fig. 2f). Deficiencies in these structures are characteristic of human genodermatoses with abnormal skin barrier function, including ichthyosis vulgaris and harlequin ichthyosis. No regions of normal keratohyalin granule formation were observed in TINCR-deficient epidermis, and the number of lamellar bodies in the stratum granulosum of TINCR-deficient human epidermal tissue was reduced by 81.4%. TINCR is thus required for the induction of genes that form the cellular structures that mediate differentiation-associated epidermal barrier formation.

To determine the mechanisms of TINCR action, we developed two assays to analyse the *TINCR* RNA and protein interactome (Supplementary Fig. 2d). Given its cytoplasmic location (Supplementary Fig. 1e), TINCR control of epidermal barrier genes may occur at the post-transcriptional level through direct association with target mRNAs. To test this, we developed RNA interactome analysis, followed by deep sequencing (RIA-Seq) (Supplementary Fig. 2d, left). Thirty-eight biotinylated DNA probes (Supplementary Table 3) were designed in even- and odd-numbered pools (Supplementary Fig. 2e).

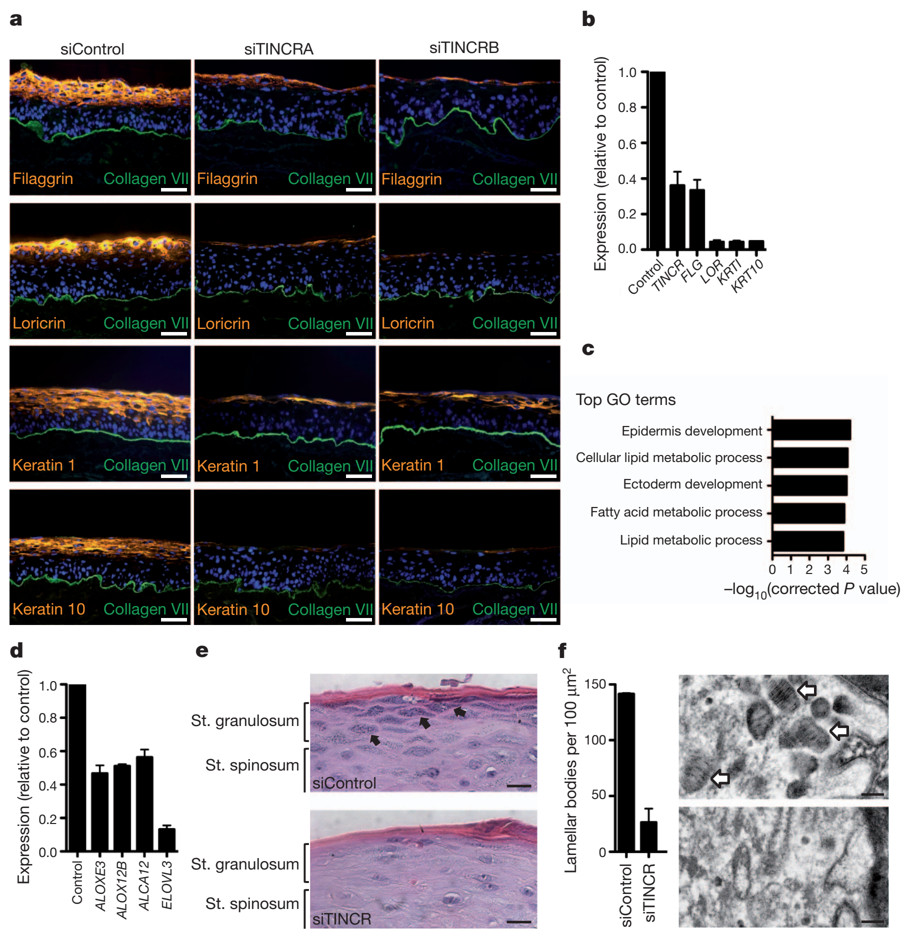


Figure 2 | TINCR regulates epidermal differentiation genes involved in barrier formation. **a**, Loss of differentiation proteins in TINCR-depleted organotypic human epidermis by independent *TINCR* siRNAs (siTINCRA and siTINCRB) versus scrambled control (siControl); nuclei stained blue (Hoechst 33342). Scale bars, 50 μm. **b**, mRNA expression in TINCR-deficient tissue versus control; duplicate biological replicates for duplicate independent TINCR siRNAs. Error bars are s.d., $n = 6$. **c**, GO terms significantly enriched in the TINCR-depleted gene subset. **d**, mRNA expression of lipid barrier synthesis genes in TINCR-depleted tissue. Error bars denote s.d., $n = 4$. **e**, Loss of protein-rich keratohyalin granules (arrows in control) in TINCR-deficient organotypic human epidermis. St, stratum. Scale bars, 10 μm. **f**, Loss of normal lipid-containing lamellar bodies (arrows in control, top image) in TINCR-depleted tissue (bottom image) ($n = 3$). Scale bars, 100 nm.

These two pools were used separately in a multiplex fashion for pull-down of endogenous TINCR and associated RNAs in differentiated keratinocytes (Supplementary Fig. 2d, left), similar to recent approaches to single-molecule RNA FISH[17] and chromatin isolation by RNA purification[18]. To discover transcripts enriched by TINCR pull-down, a 100-base-pair (bp) sliding window compared the even and odd signal to input across the human transcriptome (Supplementary Fig. 3a), resulting in the discovery of 3,602 enriched sites. GO analysis of TINCR-interacting genes showed enrichment of differentiation-associated genes (Fig. 3a). TINCR binding is enriched in mRNAs that are downregulated following TINCR knockdown ($P = 3.97 \times 10^{-7}$). RIA-Seq results were confirmed by RNA interactome analysis and qRT–PCR (Supplementary Fig. 3b). These data are consistent with potential TINCR action by post-transcriptional differentiation gene regulation.

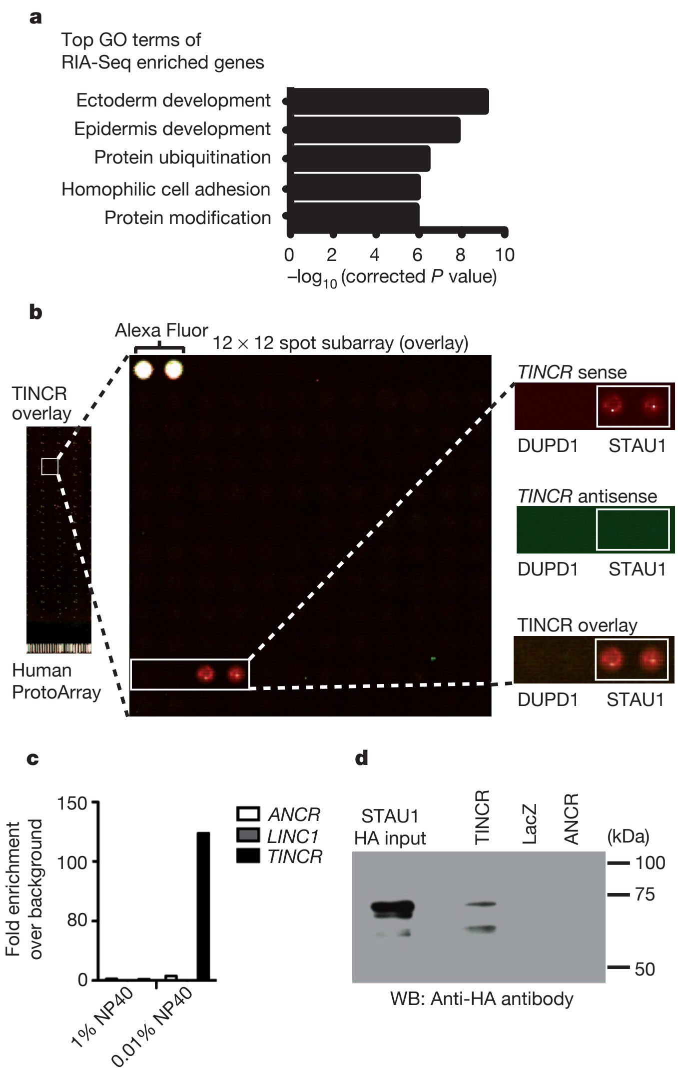


Figure 3 | TINCR interacts with differentiation mRNAs and STAU1 protein. **a**, Enriched GO terms in TINCR-interacting genes detected by RIA-Seq. **b**, Protein microarray analysis detects *TINCR* RNA binding to STAU1 protein. Human recombinant protein microarray spotted with approximately 9,400 proteins (left); enlarged 144 protein spot subarray (middle) demonstrating strand-specific binding of TINCR sense strand to STAU1 protein (right); DUPD1 protein negative control is shown. Alexa-Fluor-647-labelled rabbit anti-mouse IgG in the top left corner of each subarray. **c**, STAU1 protein immunoprecipitation pulls down *TINCR* RNA. *ANCR* and *LINC1* (also known as *XIST*) represent lncRNA controls. **d**, Streptavidin precipitation of *in vitro* synthesized biotinylated *TINCR* RNA pulls down STAU1 protein. HA, haemagglutinin; WB, western blot.

LncRNAs can act together with specific proteins[3,4,19–21]. To identify TINCR-binding proteins of relevance to epidermal differentiation control, we developed human protein microarray analysis (Supplementary Fig. 2d, right). *TINCR* sense and antisense RNA were transcribed with Cy5 and independently hybridized to a protein microarray containing approximately 9,400 recombinant human proteins (Human ProtoArray). STAU1 protein displayed the strongest *TINCR* RNA binding (Fig. 3b and Supplementary Fig. 3c). Reciprocal binding was confirmed by two-way ribonucleoprotein complex pull-down experiments (Fig. 3c, d). STAU1 is a known RNA-binding protein[22–24] first identified in *Drosophila* as a mediator of RNA localization in oocytes[25]; however, a role for STAU1 in epidermal differentiation has not been described. Similar to TINCR loss, STAU1 deficiency phenocopied impaired differentiation of epidermal tissue (Fig. 4a and Supplementary Fig. 3d). Transcript profiling of STAU1-deficient epidermis showed significant overlap of STAU1- and TINCR-regulated genes (42.5% overlap for siSTAU1 (short interfering RNA (siRNA) against *STAU1*), 47.8% for siTINCR, $P = 1.24 \times 10^{-222}$) (Fig. 4b, Supplementary Fig. 3e and Supplementary Table 4). Gene set enrichment analysis (GSEA)[26] of siSTAU1 as well as siTINCR gene sets showed marked overlap with the keratinocyte differentiation signature published previously[7] (Supplementary Fig. 3f, g) indicating that TINCR together with STAU1 is required for epidermal differentiation.

To study TINCR interaction with differentiation mRNAs, the top 1,500 TINCR enriched sites detected by RIA-Seq were subjected to a *de novo* motif search. This identified a 25-nucleotide motif that was strongly enriched in TINCR-interacting mRNAs and also repeated within TINCR itself, termed the TINCR box (Fig. 4c, Supplementary Fig. 4 and Supplementary Table 5). A reverse search for the TINCR box in all TINCR-enriched sites using the find individual motifs occurrences (FIMO) algorithm yielded sequence similarity ($P < 10 \times 10^{-4}$) for 96.3% of the 3,602 sites present. By contrast, a reverse search of the TINCR box in the least TINCR-enriched sites showed sequence similarity ($P < 1 \times 10^{-4}$) for 11.3% of analysed 2,567 sites, indicating strong enrichment of the TINCR box motif in TINCR-interacting mRNAs. To investigate whether the TINCR box is also relevant for interaction with TINCR- and STAU1-regulated mRNAs, TINCR-binding transcripts were overlapped with those showing differential expression in TINCR- and STAU1-deficient epidermis. Reverse motif search with enriched sequences of the resulting 31 genes (Supplementary Fig. 5a) also showed high occurrence of the TINCR box (Fig. 4c). Of note, 3 out of those 31 mRNAs were previously shown to interact with STAU1 protein in a kidney cell line[27]. Analysis of TINCR secondary structure in solution by selective hydroxyl acylation analyzed by primer extension (SHAPE)[28] showed that TINCR boxes of three, six, seven and ten reside in open RNA structure conformations (Supplementary Fig. 6a, b) that may be amenable to TINCR-target interactions. By contrast, analysis of microRNA (miRNA) seed sequences showed no enrichment for any miRNA seed in mRNAs affected by TINCR knockdown, arguing against a competing endogenous RNA (ceRNA) mechanism for TINCR (Supplementary Tables 6 and 7). STAU1 depletion did not affect TINCR subcellular localization (Supplementary Fig. 5e). To test whether the TINCR motif is required for target mRNA binding to TINCR, pull-down experiments were performed with a wild-type and TINCR-box-deficient target mRNA. Full TINCR binding to the *PGLYRP3* differentiation gene mRNA occurred with or without STAU1 protein, but was dependent on the 3′ *PGLYRP3* TINCR box (Fig. 4d, e). However, the existence of TINCR- and STAU1-regulated differentiation gene mRNAs that don't show direct interaction with TINCR (Supplementary Table 8) suggests the potential existence of other mechanisms of indirect target regulation in this setting.

Recent work showed that STAU1-binding sites could be created by imperfect base pairing between an ALU element of an mRNA target of STAU1-mediated decay (SMD) and another ALU sequence in a half-STAU1-binding site lncRNA; STAU1 binding to this sequence leads to

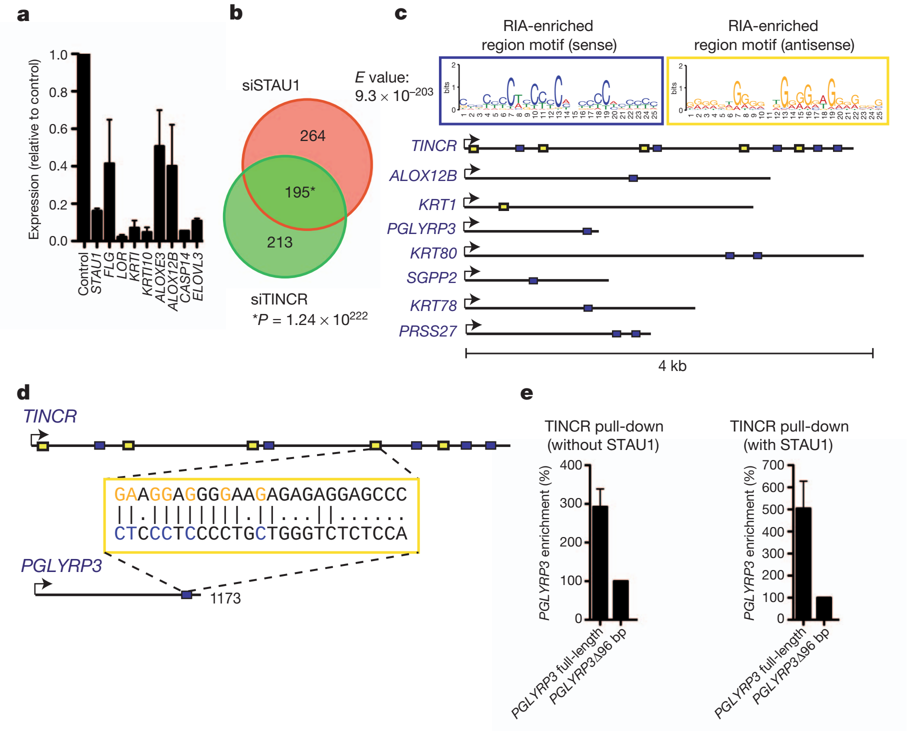


Figure 4 | Differentiation regulation by *TINCR* RNA and STAU1 protein. **a**, Diminished expression of TINCR-regulated genes in STAU1-depleted organotypic tissue in independent biological replicates for duplicate independent STAU1 siRNA treatments. Error bars are s.d., $n = 3$. **b**, Overlap of 672 genes regulated by TINCR and STAU1 **c**, Location of the TINCR box motif in *TINCR* as well as selected *TINCR*-associated differentiation mRNAs. RIA, RNA interactome analysis. **d**, TINCR motif base-pairing between the *TINCR* transcript and *PGLYRP3* differentiation gene mRNA. **e**, Biotinylated *TINCR* RNA with or without STAU1 protein pulls down the full-length 1,077-bp *PGLYRP3* mRNA at higher efficiency than TINCR box-depleted *PGLYRP3* (*PGLYRP3*Δ96 bp). Error bars denote s.d. and are too small to be visible for the *PGLYRP3*Δ96 samples; $n = 3$.

mRNA degradation in a UPF1/2-dependent manner[23]. The lack of change in TINCR transcript levels in STAU1-depleted epidermis (Supplementary Fig. 5b) and the direct binding of TINCR to STAU1 without other RNAs (Fig. 3b–d and Supplementary Fig. 3c) indicate that TINCR is neither a direct degradation target of STAU1 nor a half-STAU1-binding site RNA. Moreover, a reverse search found that only 142 of the 3,602 enriched TINCR target sites contain an ALU element, 28 of which showed enrichment for the TINCR motif ($P < 10 \times 10^{-4}$), suggesting that most TINCR–mRNA interactions occur independently of ALU elements. To test whether TINCR–STAU1 acts by UPF1/2-dependent SMD, we generated UPF1-deficient as well as UPF1/2 double-deficient human epidermal tissue. UPF depletion failed to alter epidermal differentiation substantially (Supplementary Fig. 5c, d), indicating that siTINCR differentiation defects are not due to disrupting SMD and thus occur through a previously uncharacterized mechanism. To explore this, differentiation mRNA stability was assessed as a function of TINCR and STAU1 for the two TINCR box motif-containing differentiation mRNA, *KRT80*, in differentiating keratinocytes using actinomycin D. Although *KRT80* mRNA decreased by 24.6% after 2 h in control differentiating keratinocytes, in siTINCR/siSTAU1 double-deficient keratinocytes it decreased by 91.3% (Supplementary Fig. 7), consistent with differentiation mRNA stabilization by TINCR and STAU1.

Among lncRNAs regulators[3,29,30], TINCR acts in somatic differentiation post-development. TINCR may control differentiation mRNA abundance post-transcriptionally, as indicated by TINCR cytoplasmic localization, TINCR binding to differentiation mRNAs and the effect of TINCR on differentiation mRNA stability. Binding of TINCR to interacting mRNAs occurs through the TINCR box motif. Loss of the TINCR-associated cytoplasmic protein STAU1 resembles TINCR loss and demonstrates a new UPF1/2-independent role for STAU1 in differentiation.

METHODS SUMMARY

Primary human keratinocytes were isolated from freshly discarded surgical skin samples and grown in a 1:1 mixture of KSF-M (Gibco) and Medium 154 for keratinocytes (Gibco), supplemented with epidermal growth factor and bovine pituitary extract. Cells were cultured at 37 °C in a humidified chamber with 5% CO_2. Keratinocyte differentiation was induced *in vitro* by the addition of 1.2 mM calcium to the media, then cells were grown in full confluence.

For siRNA-mediated gene transfer, 1×10^6 primary human keratinocytes were electroporated with 1 nmol siRNA oligonucleotide, using the human keratinocyte Nucleofector kit (Lonza).

Immunofluorescence analysis was performed with 7-μm-thick skin sections from human organotypic skin cultures, fixed in 100% methanol or acetone for 10 min followed by blocking in PBS with 10% bovine serum albumin (BSA) for 30 min. Sections were incubated with primary antibodies for 1 h. Antibodies were diluted in PBS with 1% BSA. For histological analysis, human organotypic epidermal tissue was fixed in 10% formalin (Sigma-Aldrich), embedded in paraffin, sectioned and stained with haematoxylin and eosin.

qRT-PCR was performed using the Maxima SYBR Green qPCR master mix (2×, Fermentas) and the Mx3000P (Agilent) thermocycler. Samples were run in triplicate and normalized to 18S ribosomal RNA or ribosomal protein L32 mRNA.

Full Methods and any associated references are available in the online version of the paper.

Received 27 May 2011; accepted 5 October 2012.
Published online 2 December 2012.

1. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).
2. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* **106,** 11667–11672 (2009).
3. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129,** 1311–1323 (2007).
4. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445,** 666–670 (2007).

5. Pauli, A., Rinn, J. L. & Schier, A. F. Non-coding RNAs as regulators of embryogenesis. *Nature Rev. Genet.* **12,** 136–149 (2011).
6. Wan, D. *et al.* Large-scale cDNA transfection screening for genes related to cancer development and progression. *Proc. Natl Acad. Sci. USA* **101,** 15724–15729 (2004).
7. Sen, G. L., Reuter, J. A., Webster, D. E., Zhu, L. & Khavari, P. A. DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* **463,** 563–567 (2010).
8. Truong, A. B., Kretz, M., Ridky, T. W., Kimmel, R. & Khavari, P. A. p63 regulates proliferation and differentiation of developmentally mature keratinocytes. *Genes Dev.* **20,** 3185–3197 (2006).
9. O'Driscoll, J. *et al.* A recurrent mutation in the loricrin gene underlies the ichthyotic variant of Vohwinkel syndrome. *Clin. Exp. Dermatol.* **27,** 243–246 (2002).
10. Smith, F. J. *et al.* Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nature Genet.* **38,** 337–342 (2006).
11. Virtanen, M., Smith, S. K., Gedde-Dahl, T. Jr, Vahlquist, A. & Bowden, P. E. Splice site and deletion mutations in keratin (*KRT1* and *KRT10*) genes: unusual phenotypic alterations in Scandinavian patients with epidermolytic hyperkeratosis. *J. Invest. Dermatol.* **121,** 1013–1020 (2003).
12. Elias, P. M. Stratum corneum defensive functions: an integrated view. *J. Invest. Dermatol.* **125,** 183–200 (2005).
13. Eckl, K. M. *et al.* Molecular analysis of 250 patients with autosomal recessive congenital ichthyosis: evidence for mutation hotspots in *ALOXE3* and allelic heterogeneity in *ALOX12B*. *J. Invest. Dermatol.* **129,** 1421–1428 (2009).
14. Sakai, K. *et al.* *ABCA12* is a major causative gene for non-bullous congenital ichthyosiform erythroderma. *J. Invest. Dermatol.* **129,** 2306–2309 (2009).
15. Westerberg, R. *et al.* Role for ELOVL3 and fatty acid chain length in development of hair and skin function. *J. Biol. Chem.* **279,** 5621–5629 (2004).
16. Denecker, G. *et al.* Caspase-14 protects against epidermal UVB photodamage and water loss. *Nature Cell Biol.* **9,** 666–674 (2007).
17. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5,** 877–879 (2008).
18. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **44,** 667–678 (2011).
19. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142,** 409–419 (2010).
20. Nagano, T. *et al.* The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322,** 1717–1720 (2008).
21. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329,** 689–693 (2010).
22. Dugré-Brisson, S. *et al.* Interaction of Staufen1 with the 5′ end of mRNA facilitates translation of these RNAs. *Nucleic Acids Res.* **33,** 4797–4812 (2005).
23. Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **470,** 284–288 (2011).
24. Kiebler, M. A. *et al.* The mammalian staufen protein localizes to the somatodendritic domain of cultured hippocampal neurons: implications for its involvement in mRNA transport. *J. Neurosci.* **19,** 288–297 (1999).
25. St Johnston, D., Beuchle, D. & Nusslein-Volhard, C. Staufen, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell* **66,** 51–63 (1991).
26. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
27. Furic, L., Maher-Laporte, M. & DesGroseillers, L. A genome-wide approach identifies distinct but overlapping subsets of cellular mRNAs associated with Staufen1- and Staufen2-containing ribonucleoprotein complexes. *RNA* **14,** 324–335 (2008).
28. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127,** 4223–4231 (2005).
29. Bond, A. M. *et al.* Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nature Neurosci.* **12,** 1020–1027 (2009).
30. Young, T. L., Matsuda, T. & Cepko, C. L. The noncoding RNA *Taurine Upregulated Gene 1* is required for differentiation of the murine retina. *Curr. Biol.* **15,** 501–512 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by US Veterans Affairs Office of Research and Development funding to P.A.K. and National Institutes of Health (National Institute of Arthritis and Musculoskeletal and Skin Diseases) grant AR49737 to P.A.K., and by NIH R01-HG004361 and California Institute for Regenerative Medicine to H.Y.C. H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

Author Contributions M.K. designed and executed experiments, analysed data and wrote the manuscript. D.E.W., Z.S., C.C., A.Z., C.S.L., R.J.F., K.Q., J.C., D.J., G.X.Y.Z., G.E.K., A.F.G., R.C.S., R.A.F. and S.A. executed experiments, analysed data and contributed to design of experimentation. A.R. and J.L.R. helped design experiments and analysed data. P.A.K. and H.Y.C. designed experiments, analysed data and wrote the manuscript.

Author Information Sequence and array data are deposited in the Gene Expression Omnibus database under the accession number GSE35468. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.Y.C. (howchang@stanford.edu) or P.A.K. (khavari@stanford.edu).

METHODS

Tissue culture. Primary human keratinocytes were isolated from freshly discarded skin surgical samples and grown in a 1:1 mixture of KSF-M (Gibco) and Medium 154 for keratinocytes (Gibco), supplemented with epidermal growth factor and bovine pituitary extract. Cells were cultured at 37 °C in a humidified chamber with 5% CO_2. Keratinocyte differentiation was induced *in vitro* by the addition of 1.2 mM calcium to the media, then cells were grown in full confluence for up to 8 days.

Gene transfer. All siRNA oligonucleotide duplexes used in this work were synthesized by Dharmacon. One-million primary human keratinocytes were electroporated with 1 nmol siRNA oligonucleotide, using the Amaxa human keratinocyte Nucleofector kit (Lonza) as well as Amaxa Nucleofection reagents as described previously[8]. The following siRNA oligonucleotides were used for this work: siControl (sense sequence): 5′-GUAGAUUCAUAUUGUAAGGUU-3′; siTINCRA (sense sequence): 5′-GCAUGAAGUAGCAGGUAUUUU-3′; siTINCRB (sense sequence): 5′-GAUCCCGAGUGAGUCAGAAUU-3′; siSTAU1A (target sequence): 5′-GCAGGGAGUUUGUGAUGCA-3′; siSTAU1B (target sequence): 5′-CGAGUAAAGCCUAGAAUCA-3′; siUPF1A (target sequence): 5′-CAGCGGAUCGUGUGAAGAA-3′; siUPF1B (target sequence): 5′-CAAGGUCCCUGAUAAUUAU-3′.

Organotypic human epidermal tissue. For the generation of organotypic human epidermis, primary human keratinocytes were nucleofected with siRNA oligonucleotides and cultured for 12–36 h. Four-hundred-thousand nucleofected keratinocytes were seeded onto devitalized dermis and raised to the air–liquid interface to initiate stratification and differentiation of the epidermis culture as described previously[7,8].

Immunofluorescence and tissue analysis. Seven-micrometre-thick skin sections from human organotypic skin cultures were fixed in 100% methanol or acetone for 10 min followed by blocking in PBS with 10% bovine serum albumin (BSA) for 30 min. Sections were incubated with primary antibodies for 1 h. Primary antibodies were diluted in PBS with 1% BSA and include collagen type VII (Calbiochem, 234192 for rabbit; Millipore, MAB2500 for mouse) at 1:200 dilution, filaggrin (Biomedical Technologies, BT-576) at 1:500 dilution, keratin 1 (Covance, PRB-149P) at 1:2,000 dilution, keratin 10 (Neomarkers, MS-611-P) at 1:200 dilution, loricrin (Covance, PRB-145P) at 1:200 dilution, and transglutaminase1 (Biomedical Technology, BT-621) at 1:100 dilution. The secondary antibodies used were Alexa-555-conjugated goat anti-mouse and goat anti-rabbit IgG (Molecular Probes, 1:300 dilution), and Alexa-488-conjugated goat anti-rabbit and goat anti-mouse IgG (Molecular Probes, 1:300 dilution). The nuclear dye Hoechst 33342 (Molecular Probes) was used at 1:1,000 dilution. For histological analysis, human organotypic epidermal tissue was fixed in 10% formalin (Sigma-Aldrich), embedded in paraffin, sectioned and stained with haematoxylin and eosin.

Transmission electron microscopy. Organotypic human epidermal tissue was fixed in 2% paraformaldehyde, 2% glutaraldehyde, 0.1 M cacodylate buffer, pH 7.3, 0.06% $CaCl_2$, post-fixed with ruthenium tetroxide (Polysciences) for 45 min and stained for 15 min in 1:1 saturated uranyl acetate. After three washes in water, samples were dehydrated and infiltrated with EMbed-812 resin (EMS) mixed 1:1 with propylene oxide for 2 h followed by two parts EMbed-812 to one part propylene oxide overnight. The samples were then placed into EMbed-812 for 2–4 h, transferred into moulds with fresh resin, oriented and incubated at 65 °C overnight. Ultrathin sections (80 nm) were picked up on formvar/carbon-coated slot copper grids and examined with the JEOL JEM-1400 TEM at 80 kV. Photos were taken using a Gatan Orius digital camera.

qRT–PCR analysis. Total RNA from cells was isolated with Trizol reagent (Invitrogen) and genomic DNA removed using the TurboDNase kit (Ambion) and quantified by NanoDrop. Total RNA from organotypic skin cultures was isolated with the RNeasy Plus mini kit (Qiagen) according to the manufacturer's instructions. One-microgram of total RNA was reverse transcribed with the iScript cDNA synthesis kit (Biorad). qRT–PCR was performed using the Maxima SYBR Green qPCR master mix (2×, Fermentas) and the Stratagene Mx3000P (Agilent Technologies) thermocycler. Samples were run in triplicate and normalized to 18S rRNA, *GAPDH* or ribosomal protein L32. The following primer sequences were used: 18S forward: 5′-GCAATTATTCCCCATGAACG-3′, reverse: 5′-GGCCTCACTAAACCATCCAA-3′; L32 forward: 5′-AGGCATTGACAACAGGGTTC-3′, reverse: 5′-GTTGCACATCAGCAGCACTT-3′; *GAPDH* forward: 5′-GAAGAGAGAGACCCTCACTGCTG-3′, reverse: 5′-ACTGTGAGGAGGGGAGATTCAGT-3′; *FLG* forward: 5′-AAAGAGCTGAAGGAACTTCTGG-3′, reverse: 5′-AACCATATCTGGGTCATCTGG-3′; *KRT1* forward: 5′-TGAGCTGAATCGTGTGATCC-3′, reverse: 5′-CCAGGTCATTCAGCTTGTTC-3′; *KRT10* forward: 5′-GCAAATTGAGAGCCTGACTG-3′, reverse: 5′-CAGTGGACACATTTCGAAGG-3′; *LOR* forward: 5′-CTCTGTCTGCGGCTACTCTG-3′, reverse: 5′-CACGAGGTCTGAGTGACCTG-3′; *STAU1* forward: 5′-ATGGTATCGGCAAGGATGTG-3′, reverse: 5′-AGACATTGGTCCGTTTCCTG-3′; *ALOX12B* forward: 5′-AGACTGCAATTCCGGATCAC-3′, reverse: 5′-TGTGGAATGCACTGGAGAAG-3′; *ALOXE3* forward: 5′-GAGCAAAAATCTCGCCAGTC-3′, reverse: 5′-GGGCTTTGTCTCAGAAATCG-3′; *ABCA12* forward: 5′-AACAGTCCAAAGCCATCCAG-3′, reverse: 5′-GAGCAGCAGCAATTTCACAG-3′; *CASP14* forward: 5′-TTCCGAAGAAGACCTGGATG-3′, reverse: 5′-TGGGGTCTCTTTTCATGGTG-3′; *ELOVL3* forward: 5′-TTCGAGGAGTATTGGGCAAC-3′, reverse: 5′-GAAGATTGCAAGGCAGAAGG-3′; *TINCR* forward: 5′-TGTGGCCCAAACTCAGGGATACAT-3′, reverse: 5′-AGATGACAGTGGCTGGAGTTGTCA-3′; *ANCR* forward: 5′-GCCACTATGTAGCGGGTTTC-3′, reverse: 5′-ACCTGCGCTAAGAACTGAGG-3′; *LINC1* forward: 5′-TTCTGGATGCAGCCACACTTCACA-3′, reverse: 5′-TGCCAGAGGAATTCTGTTTT-3′; *UPF1* forward: 5′-ATATGCCTGCGGTACAAAGG-3′, reverse: 5′-AGCTCAATGGCGATCTCATC-3′; *HIST1H2BG* forward: 5′-ACAAGCGCTCGACCATTACCT-3′, reverse: 5′-TGGTGACAGCCTTGGTACCTTC-3′; *PGLYRP3* forward: 5′-GCCAGGCAGTCTCATTTACC-3′, reverse: 5′-AGAGAAGCCAGCATCACCTC-3′; *PRSS27* forward: 5′-AGTTCATGCCCGTCTCAAAG-3′, reverse: 5′-GCCCTTCACCAATTACATCCT-3′.

Northern blot analysis. Total RNA from keratinocytes was isolated with Trizol reagent (Invitrogen). A TINCR-specific, radioactive DNA probe with a length of 581 bp was generated using [α-^{32}P]dCTP (Perkin Elmer) and the Megaprime DNA labelling system (GE Healthcare). Hybridization was performed using QuickHyb (Agilent), following the manufacturer's instructions.

Single–molecule RNA FISH. Keratinocytes were grown on chambered cover glasses (Lab-Tek) and fixed in 3.7% glutaraldehyde in PBS. Forty fluorophore-linked antisense DNA probes targeting the *TINCR* full-length sequence were designed using the online designer at http://www.singlemoleculefish.com. Probes mapping to homologous or repeat sequences were excluded. *In situ* hybridization and imaging was performed as described previously[17].

Full transcriptome RNA sequencing. RNA sequencing libraries were prepared with the mRNA Seq sample prep kit (Illumina), following the manufacturer's instructions. In brief, total RNA was isolated from differentiated (days 3 and 6) and progenitor human keratinocyte populations, and mRNA was extracted by polyA selection from 1–2 μg total RNA per sample. PolyA-selected RNA was fragmented and randomly primed for reverse transcription, followed by second-strand complementary DNA synthesis, end repair, adenylation of 3′ ends, adaptor ligation and PCR amplification of approximately 300-bp cDNA fragments. High-throughput full transcriptome sequencing was undertaken using the Illumina paired-end HiSeq platform with a read length of 101 nucleotides. Reads were aligned to the human reference sequence NCBI Build 36.1/hg18 with the TopHat algorithm[31], resulting in between 100 million and 110 million mapped reads per sample. Differential expression analysis was performed with Cuffdiff, using human RefSeq transcripts as a reference transcriptome. Annotated RefSeq non-coding RNAs expressed during calcium-induced differentiation at an FPKM > 5 in at least one time point were selected based on greater than twofold change with false discovery rate (FDR)-adjusted $P < 0.05$ and included in the cluster heat map. Non-coding RNAs that overlapped with protein-coding genes were excluded.

Gene expression profiling. Microarray analysis was performed on biological duplicate samples. Labelling of cDNA derived from control and TINCR-depleted human organotypic epidermis was done using the MessageAmp III RNA amplification kit (Ambion), following the manufacturer's instructions. Hybridization to Affymetrix GeneChip HG-U133 Plus 2.0 arrays was carried out at Stanford's Protein and Nucleic Acid Facility. Gene expression analysis was performed as described previously[7].

RNA interactome analysis. Antisense DNA probes were designed targeting the *TINCR* full-length sequence using the online designer at http://www.singlemoleculefish.com. Probes mapping to homologous or repeat sequences were excluded. Thirty-eight probes were generated and split into two sets based on their relative positions along the TINCR sequence such that even-numbered and odd-numbered probes were separately pooled, similar in design to other RNA-binding approaches taken successfully in single-molecule RNA FISH[17]. All probes were biotinylated at the 3′ end with an 18-carbon spacer arm (Protein and Nucleic Acid Facility, Stanford University). Cells were grown in differentiation conditions and rinsed once with room-temperature PBS, followed by fixation with 1% glutaraldehyde in PBS for 10 min at room temperature. Crosslinking was then quenched with 0.125 M glycine for 5 min. Cells were rinsed again with PBS, scraped into Falcon tubes, and pelleted at 1,500*g*. Cell pellets were snap frozen in liquid nitrogen and could be stored at −80 °C. To prepare lysates, cell pellets were quickly thawed in a 37 °C water bath and resuspended in cell lysis buffer (50 mM Tris, pH 7.0, 10 mM EDTA, 1% SDS, and added just before use: dithithreitol (DTT), phenylmethylsulphonyl fluoride (PMSF), protease inhibitor and Superase-In) at 100 mg ml^{-1} on ice for 10 min, and sonicated using Bioruptor (Diagenode) until

lysates were completely solubilized. RNA was in the size range of 100 to 500 nucleotides. Cell lysate was diluted in double the volume of hybridization buffer (500 mM NaCl, 1% SDS, 100 mM Tris, pH 7.0, 10 mM EDTA, 15% formamide, and added just before use: DTT, PMSF, protease inhibitor, and Superase-In. Probes (100 pmol) were added to 3 ml of diluted lysate, which was mixed by end-to-end rotation at 37 °C overnight. Streptavidin-magnetic C1 beads were washed three times in cell lysis buffer, 1 mg (100 μl) of beads was added to hybridization reaction per 100 pmol of probes, and the whole reaction was mixed for 30 min at 37 °C. Beads–biotin-probes–RNA adducts were captured by magnets (Invitrogen) and washed five times with a wash buffer volume equivalent to ten times the volume of the beads (2×SSC, 0.5% SDS, fresh PMSF added). After the last wash, buffer was removed carefully. For RNA elution, beads were resuspended in 200 μl RNA proteinase K buffer (100 mM NaCl, 10 mM Tris, pH 7.0, 1 mM EDTA, 0.5% SDS) and 1 mg ml^{-1} proteinase K (Ambion). After incubation at 50 °C for 45 min, followed by boiling for 10 min, RNA was isolated using Trizol reagent. Eluted RNA was subject to DNase treatment (TURBO DNase kit, Ambion) followed by qRT–PCR for the detection of enriched transcripts. For TINCR box removal, full-length *PGLYRP3* (1,077-bp length) was cut with the restriction endonuclease CspCI, resulting in loss of a 96-bp fragment at the 3′ end of the transcript. *LacZ* RNA was added as control. qRT–PCR data was normalized to *TINCR* and *LacZ*. For high-throughput sequencing, RNA was first converted into cDNA using the Ovation V2 kit (Nugen). cDNA samples were then fragmented to an average of 130 bp by Covaris sonicator, and 200 ng of shortened cDNA samples was converted into libraries compatible with Illumina Genome Analyzer IIx using the NEBNext ChIP-seq Library prep set (NEB). Approximately 22 million to 32 million reads were mapped per sample and used for downstream analysis.

RIA-Seq data analysis. Sequencing reads were mapped to NCBI build 36.1/hg18 using TopHat[31] with default options. Search for enriched peaks in the 'even' and 'odd' samples compared to input control was performed by scanning each gene using 100-nucleotide sliding windows. The number of reads overlapping with each sliding window was calculated using the 'coverageBed' function from BEDTools for even, odd, and input samples[32]. Signals in even, odd and input samples were defined as normalized read counts to the total number of aligned reads in each sample. Only sliding windows that have at least one read in every sample were included for downstream analysis. Enrichment score of each sliding window was defined as:

$$\mathrm{EScore}_i = \frac{\min(\mathrm{even\_signal}_i, \mathrm{odd\_signal}_i)}{\mathrm{input\_signal}_i}$$

The average $\log_2$(EScore) and standard deviation of $\log_2$(EScore) were also calculated and noted here as mean and s.d. A sliding window was defined as an enriched site if its $\log_2$(EScore) $>$ mean + 1.5 s.d. We discovered 3,602 enriched sliding windows, which cover 1,852 RefSeq genes (defined to be enriched genes). Enriched sliding windows were ranked by EScore and sequences of the top 1,500 sliding windows were extracted and used for *de novo* motif search in MEME[33]. Sequences of genes that were enriched in RIA-Seq and also differentially expressed in siTINCR and siSTAU (significance analysis of microarrays FDR < 0.05 and greater than twofold change in signal intensity in siControl versus siTINCR or siSTAU1 samples) were also extracted for reverse motif search using the FIMO algorithm[33]. GO terms associated with enriched genes were calculated using DAVID[34].

ProtoArray hybridization. The sense and antisense *TINCR* RNAs were *in vitro* transcribed from the pBluescriptR-TINCR vector (Open Biosystems) using T7 and T3 polymerases (Promega) in accordance with the manufacturer's instructions. RNA labelling for microarray hybridizations was performed using the Label IT μArray Cy5 labelling kit (Mirus) to achieve the labelling efficiency of roughly 3 pmol Cy5 dye per microgram of RNA. Human Protein Microarrays v5.0 (Invitrogen) were independently hybridized with 10 pmol sense and antisense strands of *TINCR* RNA in 40 mM Tris-HCl, pH 8.0, 150 mM sodium chloride, 0.5 mM magnesium acetate, 10 μg ml^{-1} Yeast transfer RNA, 10 μg ml heparin, 1 mM DTT, 0.01% Igepal CA-630, 5% glycerol and 0.2 U μl^{-1} RNaseOUT (Invitrogen), incubated in the dark at 25 °C for 1 h and after extensive washes spin dried and scanned at 635 nm (Cy5) using the GenePix 4000B Microarray scanner (Molecular Devices). The intensity of the 635-nm signal at each spotted protein location was determined with GenePix Pro 6.1 software (Molecular Devices). To quantify RNA–protein interactions, the intensity of the 635-nm signal (F635) was divided by the local background intensity (B635) at each of the duplicate spots for a given protein. Data were filtered based on signal above background, including duplicate features with mean signal intensities greater than twofold above local background into the analysis. For the visualization process, the array images from antisense RNA hybridizations were pseudocoloured green and overlaid with the sense RNA hybridization images.

Protein immunoprecipitation with subsequent RNA detection. For *in vitro* protein–RNA complex immunoprecipitation, pcDNA3.1Hygro-STAU1-HA plasmid containing carboxy terminal 3×HA-tagged human staufen1 was *in vitro* translated using the rabbit reticulocyte lysate system (Promega) according to the manufacturer's instructions. The sense strands of full-length human non-coding RNAs (*TINCR*, *LINC1* and *ANCR*) were *in vitro* transcribed using T7 promoter. Two-hundred-and-fifty nanograms of *TINCR*, *LINC1* (325 nucleotides; chr1q21.3) or *ANCR* (855 nucelotides; chr4q12) RNA were combined with 12.5 μl *in vitro* translated STAU1–HA in immunoprecipitation buffer containing 40 mM Tris-HCl, pH 8.0, 150 mM sodium chloride, 0.5 mM magnesium acetate, 20 μg ml^{-1} heparin, 1 mM DTT, 0.01% or 1% Igepal CA-630, 5% glycerol, 0.2 U μl^{-1} RNaseOUT (Invitrogen) and protease inhibitor Complete Mini (Roche). After 1 h of incubation STAU1–HA was immunoprecipitated with 20 μl Protein G Dynabeads (Invitrogen), bound to 2.5 μg anti-HA antibody (HA.11, clone 16B12, Covance), washed extensively with immunoprecipitation buffer, and RNA from the residual mixture was extracted with phenol–chloroform followed by ethanol precipitation. The cDNA was then synthesized with iScript cDNA synthesis kit (Bio-Rad) and subjected to qRT–PCR.

***In vitro* RNA pull-down with subsequent protein detection.** For *in vitro* RNA pull-down, *TINCR*, *LacZ* and *ANCR* mRNAs were labelled with biotin-16-UTP (GE Healthcare) during *in vitro* transcription reaction according to the manufacturer's instructions. Five microlitres of *in vitro* translated STAU1–HA was incubated with 1 μg biotin–16-UTP-labelled *TINCR*, *LacZ* or *ANCR* in immunoprecipitation buffer for 30 min at 25 °C. After addition of 5 μl of MyOne Streptavidin T1 Dynabeads (Invitrogen), the mixture was incubated for a further 30 min and subjected to five wash cycles of 5 min, each using 500 μl IPB buffer. After the final wash, magnetic beads were resuspended in 12 μl protein-loading buffer, RNA-bound protein separated by SDS–PAGE and detected with anti-HA monoclonal antibody (Covance, MMS-101P) by western blot analysis. Western blot analysis was performed as described previously[8].

RNA stability assay. For analysis of RNA stability, keratinocytes were treated with actinomycin D (6 μg ml^{-1}) at day 3 of differentiation. Cells were collected at 0, 4, 12 and 20 h time points and RNA extracted using Trizol reagent (Invitrogen). Reverse transcription was performed using oligo(dT) primers and mRNA levels were measured by qRT–PCR.

Characterization of TINCR motif box secondary structure. For the acylation of TINCR in a typical *in vitro* modification protocol, 6 μg total RNA was heated in metal-free water for 2 min at 95 °C. The RNA was then flash-cooled on ice. The RNA 3× SHAPE buffer (333 mM HEPES, pH 8.0, 20 mM magnesium chloride, 333 mM sodium chloride) was added and the RNA was allowed to equilibrate at 37 °C for 10 min. To this mixture, 1 μl of 10× electrophile (130 mM, NMIA) stock, with or without dimethylsulphoxide, was added. The reaction was permitted to continue until the desired time. Reactions were extracted once with acid phenol–chloroform (pH 4.5 ± 0.2) and twice with chloroform. RNA was precipitated with 40 μl of 3 M sodium acetate buffer, pH 5.2, and 1 μl of glycogen (20 μg μl^{-1}). Pellets were washed twice with 70% ethanol and resuspended in 10 μl RNase-free water. For reverse transcription of modified RNA, [^{32}P]-end-labelled DNA primers were annealed to 3 μg of total RNA by incubating at 95 °C for 2 min followed by a step-down cooling (2 °C s^{-1}) to 4 °C. To the reaction first-strand buffer, DTT and deoxynucleotides were added. The reaction was preincubated at 52 °C for 1 min, then superscript III (2 U μl^{-1} final concentration) was added. Extensions were performed for 10 min. To the reaction, 1 μl of 4 M sodium hydroxide was added and allowed to react for 5 min. Ten microlitres of gel loading buffer II (Ambion) was then added and cDNA extensions were resolved on 8% denaturing (7 M urea) polyacrylamide gels (29:1 acrylamide:bisacrylamide, 1% TBE).

For the characterization of reverse transcription stops, cDNA extensions were visualized by phosphorimaging (STORM, Molecular Dynamics). cDNA bands were integrated with SAFA[35]. SHAPE reactivities were normalized to a scale spanning 0 to 1.5, in which 1.0 is defined as the mean intensity of highly reactive nucleotides[36]. RNA secondary structures were predicted using mFOLD software[37]. For secondary structure analysis, the following TINCR radiolabelled RT primers were used: (1) 5′-CCTTGATGTGGTAGCGCTTCCAGCGC-3′; (2) 5′-AGACAGGGCACCCAGGGCCCCAAGAGG-3′; (3) 5′-TGTCCTGGGCAAGAGCGGAAGTGCCTC-3′; (4) 5′-AAAGCAGCTCCAGCAGGTCTGCCTGCCG-3′; (5) 5′-GGGCACACAGTGGGTCTCTGGGGACAAAG-3′; (6) 5′-TACTGTTTGTTAAATGTCAAAACACCCTG-3′; (7) 5′-TTTCCTCGGTGTGGCTGTGGGACCTTAGG-3′; (8) 5′-AAGCCTATCAGGCCTGGAGCTTCCTCAAAG-3′; (9) 5′-GAAGTAGCAGGTATTGAAGCTAGG-3′; and (10) 5′-TGCTGGGAGGAGACACAGACCTCC-3′.

31. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
32. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

33. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37,** W202–W208 (2009).
34. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).
35. Das, R., Laederach, A., Pearlman, S. M., Herschlag, D. & Altman, R. B. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* **11,** 344–354 (2005).
36. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl Acad. Sci. USA* **106,** 97–102 (2009).
37. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31,** 3406–3415 (2003).

LETTER

doi:10.1038/nature11674

COUP-TFII inhibits TGF-β-induced growth barrier to promote prostate tumorigenesis

Jun Qin[1], San-Pin Wu[1], Chad J. Creighton[2], Fangyan Dai[1,3], Xin Xie[1], Chiang-Min Cheng[1], Anna Frolov[4], Gustavo Ayala[4]†, Xia Lin[3], Xin-Hua Feng[1,3], Michael M. Ittmann[4], Shaw-Jenq Tsai[1,5], Ming-Jer Tsai[1,6] & Sophia Y. Tsai[1,6]

Mutations in phosphatase and tensin homologue (PTEN) or genomic alterations in the phosphatidylinositol-3-OH kinase-signalling pathway are the most common genetic alterations reported in human prostate cancer[1–4]. However, the precise mechanism underlying how indolent tumours with PTEN alterations acquire metastatic potential remains poorly understood. Recent studies suggest that upregulation of transforming growth factor (TGF)-β signalling triggered by PTEN loss will form a growth barrier as a defence mechanism to constrain prostate cancer progression[5], underscoring that TGF-β signalling might represent a pre-invasive checkpoint to prevent PTEN-mediated prostate tumorigenesis. Here we show that COUP transcription factor II (COUP-TFII, also known as NR2F2)[6–9], a member of the nuclear receptor superfamily, serves as a key regulator to inhibit SMAD4-dependent transcription, and consequently overrides the TGF-β-dependent checkpoint for PTEN-null indolent tumours. Overexpression of COUP-TFII in the mouse prostate epithelium cooperates with PTEN deletion to augment malignant progression and produce an aggressive metastasis-prone tumour. The functional counteraction between COUP-TFII and SMAD4 is reinforced by genetically engineered mouse models in which conditional loss of SMAD4 diminishes the inhibitory effects elicited by COUP-TFII ablation. The biological significance of COUP-TFII in prostate carcinogenesis is substantiated by patient sample analysis, in which COUP-TFII expression or activity is tightly correlated with tumour recurrence and disease progression, whereas it is inversely associated with TGF-β signalling. These findings reveal that the destruction of the TGF-β-dependent barrier by COUP-TFII is crucial for the progression of PTEN-mutant prostate cancer into a life-threatening disease, and supports COUP-TFII as a potential drug target for the intervention of metastatic human prostate cancer.

COUP-TFII is known to promote tumour angiogenesis through its function in the tumour microenvironment[10,11], but its role in tumour cells remains undefined. Consistent with aberrant expression of COUP-TFII reported in various tumours[12–17], Oncomine expression analysis revealed higher COUP-TFII expression in prostate tumour cells and in metastatic prostate cancer than in primary prostate cancer[12,13] (Supplementary Fig. 1a). To assess the clinical relevance of COUP-TFII in human prostate cancer, we stained for COUP-TFII in a tumour tissue microarray (TMA) consisting of 407 patient specimens. Examination of clinical prostate specimens showed higher levels of COUP-TFII expression in prostate tumour cells in comparison to the adjacent normal prostate epithelium (Supplementary Fig. 1b). As summarized in Fig. 1a, approximately 60% of human prostate cancer specimens exhibited intermediate to intense nuclear COUP-TFII staining, whereas only 5% of normal prostate epithelial cells stained positive for COUP-TFII. Further correlation studies indicated that COUP-TFII expression in tumour cells significantly associated with pathological predictors of human prostate cancer aggressiveness (Supplementary Table 1). Most importantly, univariate analysis indicated that COUP-TFII expression levels in prostate tumour cells could serve as a predictor to stratify risk of recurrence in patients; patients bearing higher levels of COUP-TFII expression showed earlier recurrence after radical prostatectomy (Fig. 1b). Cox proportional hazards regression analyses showed that tumour recurrence was sensitive to gradual increases of COUP-TFII expression levels, as reflected by the increasing hazard ratios, from 1.5 to 1.8 and 2.5.

These results prompted us to use genetically engineered mouse models to determine the causal role of COUP-TFII in prostate tumorigenesis. As such, we engineered mice containing PB-Cre (the *Cre* gene under control of the prostate-specific probasin promoter)[18] and conditional knockout alleles of *Pten*[3] and/or *COUP-TFII* (also known as *Nr2f2*)[7], hereafter referred to as $Pten^{PC-/-}$ and $Pten^{PC-/-}$; $COUP\text{-}TFII^{PC-/-}$ mice, respectively (Supplementary Fig. 2a, b). PTEN-null mice lacking COUP-TFII exhibited a significant reduction in prostate weight (Supplementary Fig. 2c), suggesting a positive role for COUP-TFII in driving prostate tumorigenesis. Histopathological analysis of

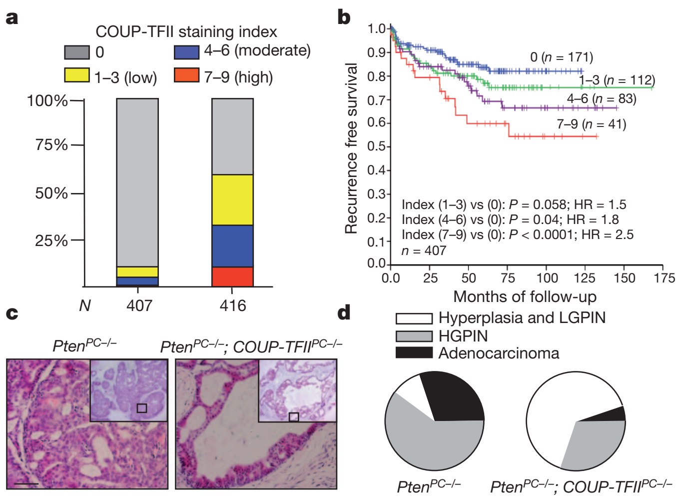


Figure 1 | COUP-TFII is crucial for prostate cancer progression in human and mice. **a**, Expression and clinical relevance of COUP-TFII are assessed in a human prostate cancer TMA. COUP-TFII staining indexes in cohorts of normal prostate epithelial cells ($n = 407$) and prostate tumours ($n = 416$) are shown as stacked columns. **b**, Kaplan–Meier plot of the recurrence after radical prostatectomy based on the COUP-TFII expression index in patients. HR, hazard ratio. **c**, Haematoxylin and eosin (H&E) staining of the anterior prostate from $Pten^{PC-/-}$ and $Pten^{PC-/-}$; $COUP\text{-}TFII^{PC-/-}$ mice at 5 months of age. Scale bar, 50 μm. **d**, Pie graphs show prostate tumour progressions in $Pten^{PC-/-}$ and $Pten^{PC-/-}$; $COUP\text{-}TFII^{PC-/-}$ mice (5 months old; $n = 10$). HGPIN, high-grade PIN; LGPIN, low-grade PIN.

[1]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. [2]Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA. [3]Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, Texas 77030, USA. [4]Department of Pathology and Immunology, Baylor College of Medicine, Houston, Texas 77030, USA. [5]Department of Physiology, College of Medicine, National Cheng Kung University, Tainan 701, Taiwan. [6]Program in Developmental Biology, Baylor College of Medicine, Houston, Texas 77030, USA.
†Present address: Department of Pathology and Laboratory Medicine, University of Texas Health Science Center Medical School, Houston, Texas 77030, USA.

the entire animal cohort showed that loss of COUP-TFII compromised prostate tumour progression, as reflected by the arrest of tumours at the hyperplastic or low-grade prostatic intraepithelial neoplasia (PIN) stage at 5 months of age, whereas tumours from $Pten^{PC-/-}$ mice already advanced to high-grade PIN or adenocarcinoma (Fig. 1c, d and Supplementary Fig. 2d). The normal prostatic histology in $COUP\text{-}TFII^{PC-/-}$ mice (Supplementary Fig. 2e) argued against a possibility that it was due to a developmental defect. Moreover, loss of COUP-TFII attenuated the proliferation advantage elicited by PTEN loss, whereas the levels of angiogenesis remained similar and thus is unlikely to have contributed to the restriction of prostate tumorigenesis (Supplementary Fig. 2f, g).

Given that COUP-TFII was further upregulated in metastatic prostate cancer, and its higher expression correlated with a worse clinical outcome, we sought to determine whether overexpression of COUP-TFII in PTEN-null tumours would exacerbate prostate tumorigenesis. As such, we generated an inducible COUP-TFII overexpression mouse, $COUP\text{-}TFII^{OE/+}$ (Fig. 2a), that allows constitutive COUP-TFII expression in the prostate epithelium. During 12 months of follow-up, $COUP\text{-}TFII^{OE/+}$ mice exhibited normal prostatic histology (Supplementary Fig. 3a), indicating that COUP-TFII by itself is not sufficient to initiate prostate neoplasia, and that cooperating oncogenic lesions are required. Indeed, overexpression of COUP-TFII in a PTEN-heterozygous background resulted in a rapid acceleration of tumour progression, with overt and high penetrance PIN starting at 4 months of age and progressing to high-grade PIN or adenocarcinoma by 12 months (Fig. 2b). These data demonstrate that COUP-TFII enables rapid disease progression from preneoplastic prostatic epithelium.

Next, we determined whether prostate-specific overexpression of COUP-TFII in a PTEN-null background would produce a metastasis-prone cancer. By 24 weeks of age, $Pten^{PC-/-}$; $COUP\text{-}TFII^{OE/+}$ mice developed highly aggressive carcinoma with an altered stroma, as evidenced by the attenuation or loss of the periglandular smooth muscle layer and androgen receptor (AR)-positive tumour cells invading into the stroma (Fig. 2c), in which subsets of invasive tumour cells exhibited epithelial–mesenchymal transition (Supplementary Fig. 3c). Molecular pathological analysis of prostate-cancer-bearing $Pten^{PC-/-}$; $COUP\text{-}TFII^{OE/+}$ mice also showed a metastatic spread of AR-positive tumour nodules to lumbar lymph nodes in 13 out of 16 cases, and lung metastasis in 4 out of 16 cases (Fig. 2d). By contrast, $Pten^{PC-/-}$ mice only showed a modest metastatic phenotype[2–4]. Even with more aggressive tumour phenotypes, angiogenesis was not altered whereas proliferation was enhanced after COUP-TFII overexpression (Fig. 2e and Supplementary Figs 3d, e and 4). Collectively, COUP-TFII overexpression enhances the full spectrum of prostate cancer development, leading to aggressive metastasis in PTEN-null tumours.

To determine how COUP-TFII affects tumorigenesis, we conducted a transcriptome comparison using COUP-TFII-depleted PC3 cells. Unbiased ingenuity pathway analysis indicated that COUP-TFII was crucial for the growth of human prostate cancer cells, an observation reinforced by anchorage-independent growth assays and cell proliferation/viability tests (MTT assay) (Supplementary Fig. 5). Further gene-set enrichment analysis and comparison with our gene expression profile to a defined TGF-β response signature[19] indicated a global enrichment of TGF-β-induced genes in the absence of COUP-TFII (Fig. 3a and Supplementary Fig. 6a). Given that TGF-β signalling is crucial for prostate cancer progression[5,20], we examined whether COUP-TFII potentiated prostate tumorigenesis through TGF-β signalling. In human prostate cancer cells, COUP-TFII served as a transcriptional inhibitor of TGF-β signalling (Supplementary Fig. 6b–d). Knockdown of COUP-TFII in PC3 cells substantially altered p21, p15 and cyclin D1 expression in response to TGF-β stimulation (Supplementary Fig. 6e). Protein levels of p21 were also increased without TGF-β stimulation, presumably owing to autocrine TGF-β signalling in PC3 cells (Supplementary Fig. 6f).

We postulated that overexpression of COUP-TFII enabled the indolent PTEN-null tumours to acquire metastatic potential through the destruction of a TGF-β-induced growth barrier. Indeed, in

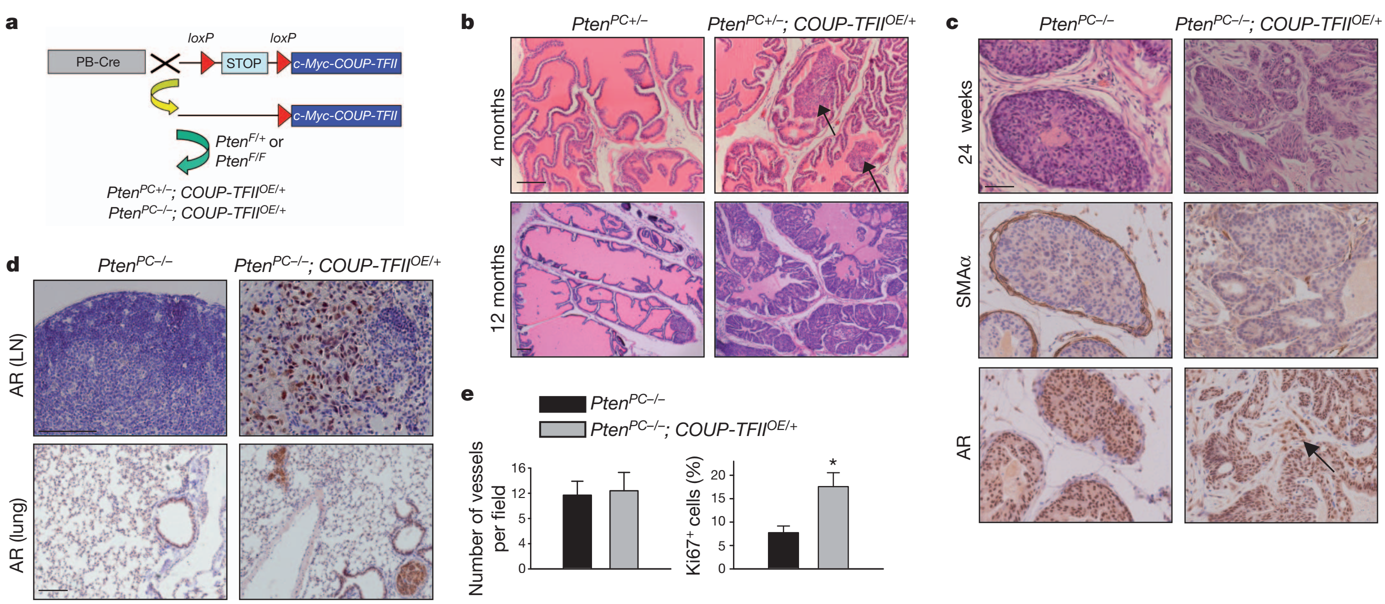
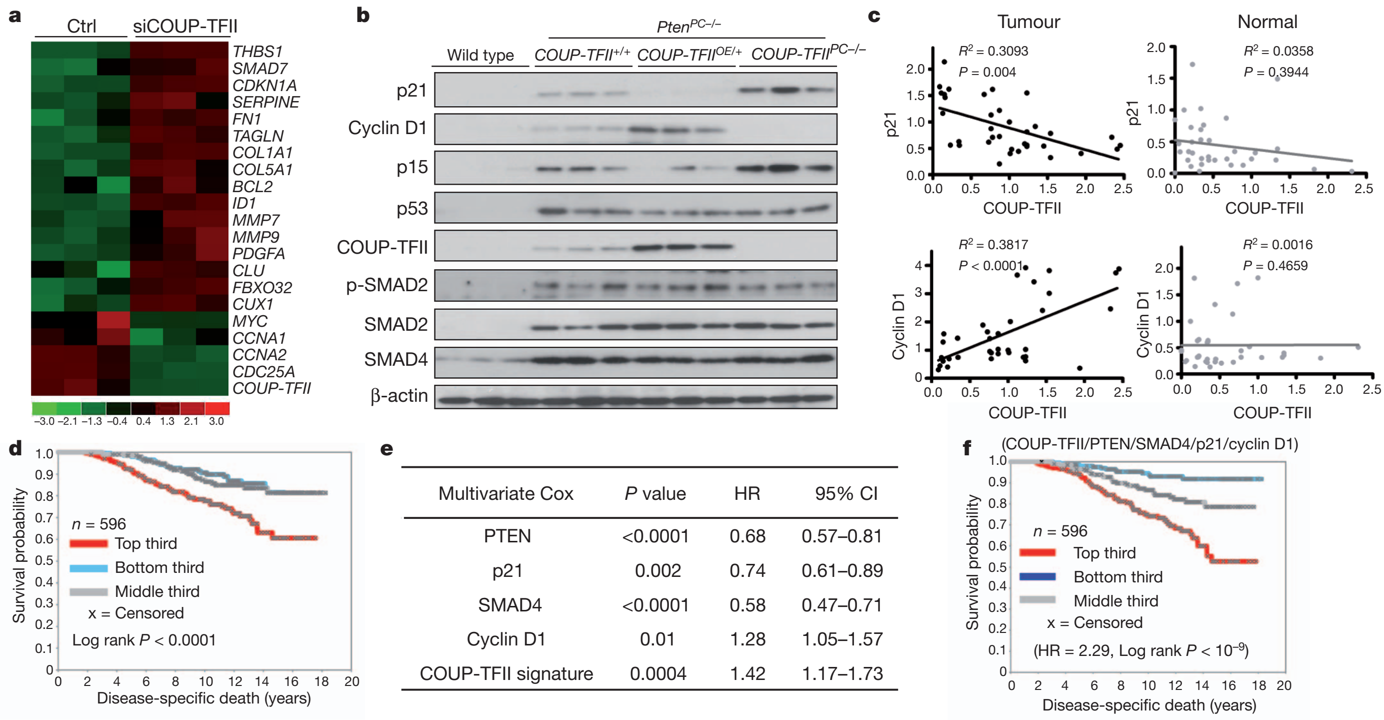


Figure 2 | Prostate-specific overexpression of COUP-TFII promotes tumorigenesis in PTEN mutated mice. **a**, Scheme of conditional overexpression of COUP-TFII (*Rosa26-lox-STOP-lox-COUP-TFII*; $COUP\text{-}TFII^{OE/+}$) in a PTEN spontaneous tumour model. $COUP\text{-}TFII^{OE/+}$ mice contain a single copy of a mini gene, consisting of a CAGGS promoter, a *loxP*-STOP-*loxP* (LSL) cassette and Myc-tagged *COUP-TFII* complementary DNA, knocked into the ROSA locus. In the presence of PB-Cre recombinase, the LSL cassette is excised, allowing COUP-TFII expression in the prostate epithelium. **b**, H&E-stained sections of representative anterior prostate at 4 and 12 months of age in $Pten^{PC+/-}$ and $Pten^{PC+/-}$; $COUP\text{-}TFII^{OE/+}$ mice. Arrows indicate PIN. **c**, H&E-stained sections of anterior prostate at 24 weeks of age in $Pten^{PC-/-}$ and $Pten^{PC-/-}$; $COUP\text{-}TFII^{OE/+}$ mice (top). Immunohistochemistry of α-smooth muscle actin (SMAα) (middle) and AR (bottom) expression. **d**, Immunohistochemical staining of AR expression in lumbar lymph nodes (LN) and lung from 6-month-old $Pten^{PC-/-}$ and $Pten^{PC-/-}$; $COUP\text{-}TFII^{OE/+}$ mice. **e**, Semi-quantitative results of Ki67 and CD31 staining of prostate tumours from 6-month-old $Pten^{PC-/-}$ and $Pten^{PC-/-}$; $COUP\text{-}TFII^{OE/+}$ mice $n = 6$ (**b**–**e**); *$P < 0.05$. Data are mean and s.e.m. Scale bars, 100 μm (**b**, **d**) and 50 μm (**c**).



Multivariate Cox	*P* value	HR	95% CI
PTEN	<0.0001	0.68	0.57–0.81
p21	0.002	0.74	0.61–0.89
SMAD4	<0.0001	0.58	0.47–0.71
Cyclin D1	0.01	1.28	1.05–1.57
COUP-TFII signature	0.0004	1.42	1.17–1.73

Figure 3 | COUP-TFII inhibits TGF-β signalling in prostate cancer cells. **a**, Heat map of microarray results shows the enrichment of TGF-β downstream target genes in COUP-TFII-depleted PC3 cells. siCOUP-TFII denotes COUP-TFII short interfering RNA (siRNA); ctrl denotes control. **b**, Western blot analysis of core regulators and downstream targets of TGF-β signalling, as well as p53, in anterior prostates from 24-week-old mice as indicated. **c**, The correlation between COUP-TFII and TGF-β signalling targets (p21 and cyclin D1) in patient prostate tumours and normal prostate counterparts are shown in regression plots ($n = 36$). **d**–**f**, Association of COUP-TFII gene signature with disease-specific death in patients with prostate cancer using data set GSE10645. **d**, Kaplan–Meier plot, with top third, middle third and bottom third of patients representing COUP-TFII activities from high to low (*P* value by log-rank test). **e**, Multivariate Cox regression analysis of COUP-TFII signature and the transcription levels of PTEN, SMAD4, p21 and cyclin D1. **f**, Kaplan–Meier plot of patients grouped using the combination of COUP-TFII and PTEN/SMAD4/cyclin D1/p21 gene sets.

Pten$^{PC-/-}$; *COUP-TFII*$^{OE/+}$ mice, the TGF-β-induced growth barrier was significantly attenuated after increased COUP-TFII expression (Fig. 3b). Overexpression of COUP-TFII profoundly neutralized oncogene-induced senescence, as judged by the reduction of senescence-associated β-galactosidase staining (Supplementary Fig. 6g), whereas ablation of COUP-TFII in PTEN-null mice enhanced the acquisition of hyperactivity of TGF-β signalling and a marked induction in cellular senescence. Furthermore, the clinical relevance between COUP-TFII levels and TGF-β signalling activities was established by a significant correlation between COUP-TFII and p21/cyclin D1 expression in prostate tumours, but not in normal prostate epithelium counterparts, in a cohort of 36 patients with highly aggressive prostate cancer (Fig. 3c).

Next, we extended our findings by querying the gene expression profiling data set GSE10645 (ref. 21) that contains 596 primary human tumours with prostate cancer-specific death information, to complement our TMA results. Using gene expression profiles obtained in PC3 cells, we found a total of 80 COUP-TFII-regulated genes represented in GSE10645 (Methods, Supplementary Fig. 7 and Supplementary Table 2); the genes were used to score each of the human tumours for COUP-TFII activity. We observed distinct differences in the time to prostate-cancer-specific death, although no differences in the time to prostate-specific antigen (PSA) recurrence, among the patients with higher versus lower COUP-TFII activity (Fig. 3d and Supplementary Table 3–5). Multivariate Cox model analysis indicated that the PTEN, SMAD4, p21, cyclin D1 and COUP-TFII signature (Methods) all independently provided predictive power to stratify patients into high- and low-risk groups (Fig. 3e). Plotting the combined transcript levels of PTEN, SMAD4, p21 and cyclin D1 together with the higher COUP-TFII signature in prostatectomies increased the predictive power from that of either gene set alone (Fig. 3f), illustrating a significant genetic cooperation between COUP-TFII and TGF-β signalling in PTEN-null human prostate cancer tumours.

Because depletion of COUP-TFII failed to affect the levels of SMAD2/3/4, the receptors or the activated form of phosphorylated-SMAD2/3 (Supplementary Fig. 8), we investigated whether COUP-TFII could interact directly with any of these SMAD proteins. Indeed, COUP-TFII strongly associated with SMAD4 in cells (Supplementary Fig. 9a–c) and in patient tumour specimens (Fig. 4a). Overexpression of mutant COUP-TFII-M2, which was unable to associate with SMAD4, failed to inhibit TGF-β signalling, indicating that the COUP-TFII–SMAD4 interaction was essential for the COUP-TFII effects on TGF-β signalling (Supplementary Fig. 9d–f). Mechanistically, our data indicated that COUP-TFII sequestered SMAD4 from binding to TGF-β-target gene promoters in cells (Supplementary Fig. 10) and in tumours containing higher levels of COUP-TFII (Fig. 4b).

As supporting evidence for the counteraction between COUP-TFII and SMAD4, knockdown of SMAD4 in PC3 cells alleviated the proliferative defects exerted by COUP-TFII depletion (Supplementary Fig. 11). Furthermore, we used a SMAD4 conditional knockout mouse[22] to examine the tumour spectrum in *Pten*$^{PC-/-}$; *COUP-TFII*$^{PC-/-}$ and *Pten*$^{PC-/-}$; *COUP-TFII*$^{OE/+}$ mice. Compared with hyperplasia or low-grade PIN lesions observed in 5-month-old *Pten*$^{PC-/-}$; *COUP-TFII*$^{PC-/-}$ mice, ablation of SMAD4 in COUP-TFII-null mice restored the full spectrum of tumour development as evidenced by the abundance of Ki67-labelled nuclei, the appearance of microinvasive tumours and the similar incidences of lymph node metastasis between *Pten*$^{PC-/-}$; *Smad4*$^{PC-/-}$ and *Pten*$^{PC-/-}$; *COUP-TFII*$^{PC-/-}$; *Smad4*$^{PC-/-}$ mice (Fig. 4c, d and Supplementary Figs 12–14). Moreover, overexpression of COUP-TFII in *Pten*$^{PC-/-}$; *Smad4*$^{PC-/-}$ mice did not further aggravate the invasive phenotypes displayed by the *Pten*$^{PC-/-}$; *Smad4*$^{PC-/-}$ mice (Fig. 4c, d and Supplementary Figs 12–14).

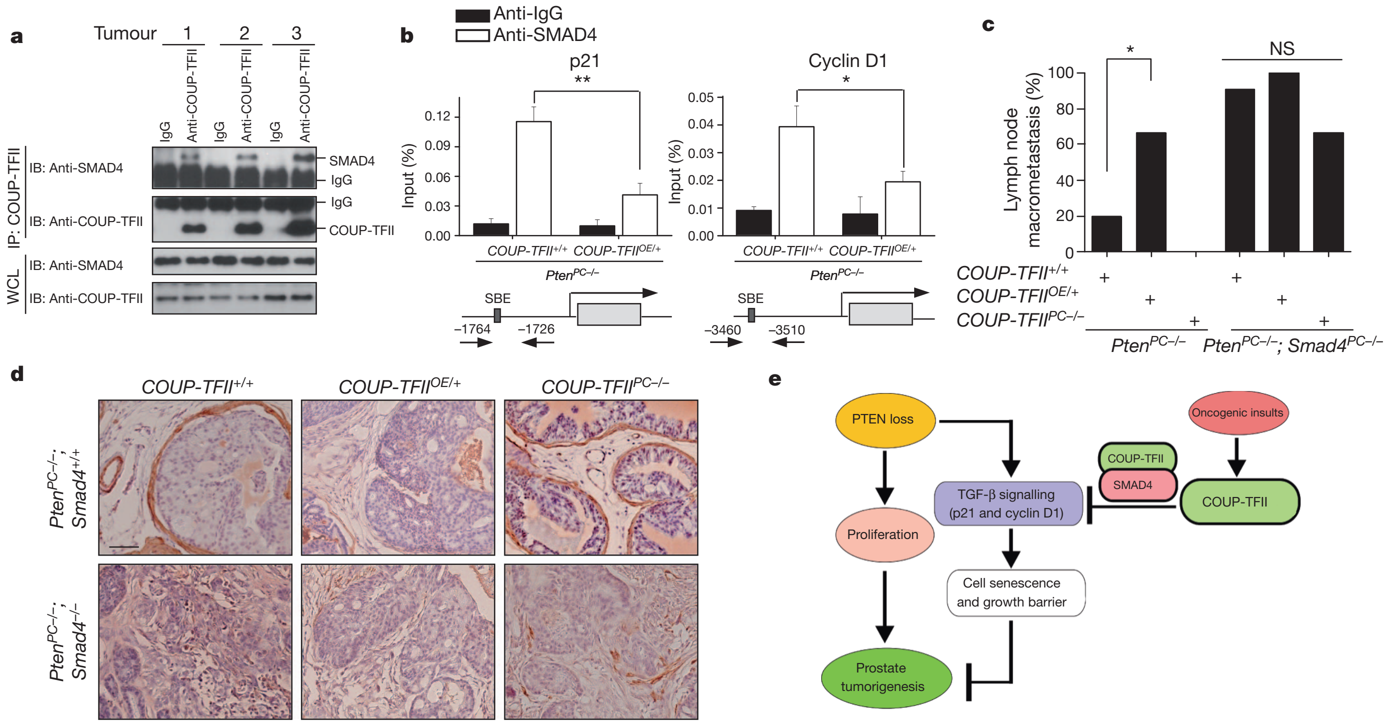


Figure 4 | COUP-TFII interacts with SMAD4 to modulate TGF-β signalling. **a**, Co-immunoprecipitation (IP) analysis of COUP-TFII and SMAD4 association in human prostate tumour specimens. IB, immunoblot. **b**, Cell lysates were prepared from prostate tumours in *Pten*$^{PC-/-}$ and *Pten*$^{PC-/-}$; *COUP-TFII*$^{OE/+}$ mice (6 months of age, $n = 3$), and subjected to chromatin immunoprecipitation analysis of SMAD4 binding at the p21 or cyclin D1 promoter. **c**, **d**, Analyses of the lymph node metastasis (**c**) and SMAα expression (**d**) in prostate tumours from *Pten*$^{PC-/-}$, *Pten*$^{PC-/-}$; *COUP-TFII*$^{OE/+}$, *Pten*$^{PC-/-}$; *COUP-TFII*$^{PC-/-}$ mice with or without SMAD4 deletion (5 months of age). **e**, Model of COUP-TFII in prostate tumorigenesis. PTEN inactivation drives prostate tumour initiation and progression. However, it also elicits the activation of TGF-β signalling that induces cellular senescence to constrain the indolent tumour from becoming aggressive. To develop metastasis-prone tumours, alternative oncogenic signals stimulate COUP-TFII expression, which counteracts the TGF-β-dependent checkpoint through direct association with SMAD4. Thus, COUP-TFII serves as a crucial regulator that counteracts the TGF-β-dependent growth barrier to enable indolent prostate cancer tumours to acquire metastatic potential. *$P < 0.05$; **$P < 0.01$. Data in **b** and **c** are mean and s.e.m.

In summary, our findings highlight a crucial role for COUP-TFII in driving full malignant progression of PTEN-null prostate tumorigenesis by counteracting TGF-β negative feedback signalling (Fig. 4e). As COUP-TFII is a member of the nuclear receptor family, the activity of which can be regulated by small molecules[23], our studies provide a potential new drug target for the intervention of metastatic human prostate cancer.

METHODS SUMMARY

Tissue microarrays used in this study were described previously[24,25]. All animal experiments were approved by the Animal Center for Comparative Medicine at Baylor College of Medicine. Detailed materials, methods and statistical analysis are provided in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 5 December 2011; accepted 10 October 2012.
Published online 28 November 2012.

1. Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18,** 11–22 (2010).
2. Trotman, L. C. *et al.* Pten dose dictates cancer progression in the prostate. *PLoS Biol.* **1,** E59 (2003).
3. Wang, S. *et al.* Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. *Cancer Cell* **4,** 209–221 (2003).
4. Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436,** 725–730 (2005).
5. Ding, Z. *et al.* SMAD4-dependent barrier constrains prostate cancer growth and metastatic progression. *Nature* **470,** 269–273 (2011).
6. Pereira, F. A., Qiu, Y., Zhou, G., Tsai, M. J. & Tsai, S. Y. The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development. *Genes Dev.* **13,** 1037–1049 (1999).
7. Takamoto, N. *et al.* COUP-TFII is essential for radial and anteroposterior patterning of the stomach. *Development* **132,** 2179–2189 (2005).
8. You, L. R. *et al.* Suppression of Notch signalling by the COUP-TFII transcription factor regulates vein identity. *Nature* **435,** 98–104 (2005).
9. Lin, F. J., Qin, J., Tang, K., Tsai, S. Y. & Tsai, M. J. Coup d'Etat: an orphan takes control. *Endocr. Rev.* **32,** 404–421 (2011).
10. Qin, J., Chen, X., Yu-Lee, L. Y., Tsai, M. J. & Tsai, S. Y. Nuclear receptor COUP-TFII controls pancreatic islet tumor angiogenesis by regulating vascular endothelial growth factor/vascular endothelial growth factor receptor-2 signaling. *Cancer Res.* **70,** 8812–8821 (2010).
11. Qin, J., Chen, X., Xie, X., Tsai, M. J. & Tsai, S. Y. COUP-TFII regulates tumor growth and metastasis by modulating tumor angiogenesis. *Proc. Natl Acad. Sci. USA* **107,** 3687–3692 (2010).
12. Dhanasekaran, S. M. *et al.* Delineation of prognostic biomarkers in prostate cancer. *Nature* **412,** 822–826 (2001).
13. Tomlins, S. A. *et al.* Integrative molecular concept modeling of prostate cancer progression. *Nature Genet.* **39,** 41–51 (2007).
14. Nagasaki, S. *et al.* Chicken ovalbumin upstream promoter transcription factor II in human breast carcinoma: possible regulator of lymphangiogenesis via vascular endothelial growth factor-C expression. *Cancer Sci.* **100,** 639–645 (2009).
15. Shin, S. W. *et al.* Clinical significance of chicken ovalbumin upstream promoter-transcription factor II expression in human colorectal cancer. *Oncol. Rep.* **21,** 101–106 (2009).
16. Dyrskjøt, L. *et al.* Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification. *Cancer Res.* **64,** 4040–4048 (2004).
17. Yeh, H. Y. *et al.* Identifying significant genetic regulatory networks in the prostate cancer from microarray data based on transcription factor analysis and conditional independency. *BMC Med. Genomics* **2,** 70 (2009).
18. Jin, C., McKeehan, K. & Wang, F. Transgenic mouse with high Cre recombinase activity in all prostate lobes, seminal vesicle, and ductus deferens. *Prostate* **57,** 160–164 (2003).
19. Padua, D. *et al.* TGFβ primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133,** 66–77 (2008).
20. Chen, M. *et al.* Identification of PHLPP1 as a tumor suppressor reveals the role of feedback activation in PTEN-mutant prostate cancer progression. *Cancer Cell* **20,** 173–186 (2011).

21. Nakagawa, T. *et al.* A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. *PLoS ONE* **3,** e2318 (2008).
22. Chu, G. C., Dunn, N. R., Anderson, D. C., Oxburgh, L. & Robertson, E. J. Differential requirements for Smad4 in TGFβ-dependent patterning of the early mouse embryo. *Development* **131,** 3501–3512 (2004).
23. Kruse, S. W. *et al.* Identification of COUP-TFII orphan nuclear receptor as a retinoic acid-activated receptor. *PLoS Biol.* **6,** e227 (2008).
24. Ayala, G. *et al.* High levels of phosphorylated form of Akt-1 in prostate cancer and non-neoplastic prostate tissues are strong predictors of biochemical recurrence. *Clin. Cancer Res.* **10,** 6572–6578 (2004).
25. Hodgson, M. C. *et al.* Decreased expression and androgen regulation of the tumor suppressor gene INPP4B in prostate cancer. *Cancer Res.* **71,** 572–582 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Wu for the *Pten*$^{flox/flox}$ mice, F. Wang for *ARR2PB-Cre* (PB-Cre) transgenic mice and E. J. Robertson for the *Smad4*$^{flox/flox}$ mice. We thank H. K. Graves and L.-Y. Yu-Lee for editorial assistance and S. Elledge for comments. We also thank the Baylor Microarray Core supported by the DERC Center (P30 DK079638) for the microarray analysis. This work was supported by grants from the National Institutes of Health DK62434, DK59820 (S.Y.T. and M.-J.T.), DK45641 (M.-J.T.) and HL76448 (S.Y.T.), and the Dan L. Duncan Cancer Center.

Author Contributions J.Q., M.-J.T. and S.Y.T. conceived and designed the experimental approach, performed experiments and prepared the manuscript as senior authors. C.J.C. contributed to computational analysis for gene signature analysis and statistical analysis. A.F., G.A. and M.M.I. performed TMA and pathology analyses. S.-P.W. generated *COUP-TFII*$^{OE/+}$ mice. F.D., X.X. and C.-M.C. performed and X.-H.F., X.L. and S.-J.T. supervised a specific subset of experimental design and analysis.

Author Information The microarray data have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE33182. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.T. (mtsai@bcm.edu) or S.Y.T. (stsai@bcm.edu).

METHODS

Immunohistochemistry and quantitative reverse transcriptase PCR analysis. Tissue microarrays used in this study were described previously[24,25]. Samples were procured from radical prostatectomies of 416 patients, who received no adjuvant therapy such as radiation or hormonal therapy. Other patient characteristics were as described previously[24–26]. Immunohistochemical analyses were done as described[24–26] using a specific anti-COUP-TFII antibody (Perseus Proteomics). Samples were scanned using a Bliss automated slide scanner to generate high-resolution digital images. The expression of COUP-TFII protein was scored and quantified as described previously[24,25]. In brief, the slides were digitized and nuclear COUP-TFII protein expression was scored and quantified based on a multiplicative index of the average staining intensity (0–3) and extent of staining (0–3) in the cores, yielding a 10-point staining index that ranged from 0 (no staining) to 9 (extensive, strong staining). Quantitative PCR (qPCR) was performed on RNAs isolated from patients with prostate cancer ($n = 36$) undergoing a radical prostatectomy for clinically localized prostate cancer with clinical characteristics as previously described[25]. The tissues were >70% cancerous based on histopathological analysis.

Animal experiments. All experiments were approved by the Animal Center for Comparative Medicine at Baylor College of Medicine. *COUP-TFII$^{flox/flox}$* mice[7] were previously generated by our laboratory, and *COUP-TFII$^{OE/+}$* mice were generated by our group using an approach as previously described[27]. *Pten$^{flox/flox}$* mice[3] were obtained from H. Wu. The *ARR2PB-Cre* (PB-Cre) transgenic mice[28] were from F. Wang. The *Smad4$^{flox/flox}$* mice[22] were from E. J. Robertson. PB-Cre, *COUP-TFII$^{flox/flox}$*, *COUP-TFII$^{OE/+}$* and *Pten$^{flox/flox}$* mice were backcrossed with C57BL/6 mice for at least five generations, and maintained according to the National Institutes of Health Guide for the Care and Use of Laboratory. *COUP-TFII*-floxed or *COUP-TFII$^{OE/+}$* mice were first crossed with *Pten*-floxed or *Smad4*-floxed mice. The resulting compound mice were then crossed with PB-Cre mice for conditional knockout or overexpression of COUP-TFII in the prostate epithelium.

Expression plasmids and siRNA. The full-length human *COUP-TFII* cDNA was cloned into pMSCV-puro/neo (Clontech), pcDNA5 (Flag tag) (Invitrogen) and pcDNA3.1 (Myc tag) (Invitrogen) to generate COUP-TFII expression plasmids. Overlapping PCR with primers was used to generate Flag-COUP-TFII-M1 and M2 constructs, which are inserted in the KPNI/NheI sites of pcDNA5. Expression plasmids for epitope-tagged SMAD2/3/4, glutathione *S*-transferase (GST)-SMAD4 and the SMAD-binding element (SBE)-Luc reporter were described previously[29]. *COUP-TFII*-specific short hairpin RNA (shRNA) constructs were generated previously[30]. In brief, the shRNA for *COUP-TFII* (5′-AGGTAACGTGA TTGATTCAGTATCTTA-3′ and 5′-AGCTCTTGCTTCGTCTCCC-3′) was cloned into pSUPER.retro (OligoEngine). SMART pool or single siRNA duplexes targeting *COUP-TFII*, *Smad4* and control non-target siRNA were purchased from Dharmacon.

Cell culture and reagents. PC3, VCaP, LNCaP and DU145 cells were obtained from the ATCC and cultured in DMEM or RPMI supplemented with 10% fetal bovine serum. RWPE-1 cells (ATCC) were cultured with keratinocyte serum-free medium (K-SFM; Invitrogen) with bovine pituitary extract and human recombinant EGF. Cells were transfected with siRNA duplexes (40–80 nM) using lipofectamine (Invitrogen) or Dharmacon transfection (Dharmacon) reagents according to the manufacturer's instructions. COUP-TFII retrovirus was generated and transfected as previously described[1]. In brief, cells were transduced with retrovirus and cultured with 2 mg ml^{-1} puromycin or 400 mg ml^{-1} neomycin and the selection was stopped as soon as the non-infected control cells died off. Cell proliferation assay was performed using CellTiter 96 non-radioactive cell proliferation assay (MTT) kit (Promega) assay according to the manufacturer's instructions. Results were calculated based on three independent experiments and statistical significance was determined by two sided Student's *t*-test.

Luciferase reporter assay. Reporter assays were carried out as described[11]. In brief, 20–24 h after transfection, cells were treated with TGF-β (10 ng ml^{-1}) or SB431542 compound (20 nM) for 24 h. Cells were then collected for measurement of luciferase and β-galactosidase activities. All assays were done in duplicate and all values were normalized for transfection efficiency against β-galactosidase activities. Luciferase activity was determined with the Promega luciferase assay kit. Results were quantified based on three independent experiments and statistical significance was determined by two sided Student's *t*-test. *P* values less than 0.05 were considered significant.

RNA isolation and qPCR. Total RNA was extracted using TRIzol followed by the RNeasy mini kit (Qiagen) cleanup and RQ1 RNase-free DNase set treatment (Promega) according to the manufacturer's instructions. First-strand cDNA was synthesized using 2 μg of total RNA and Superscript II (Invitrogen). TaqMan universal master mix reagents and inventoried primer/probe mixtures (Applied Biosytems) were used for the reaction. Standard curves were generated by serial dilution of a preparation of total RNA, and all messenger RNA quantities were normalized against 18S RNA using ABI ribosomal RNA control reagents. The primers/probes used in this study are as follows: COUP-TFII (Mm00772789_m1), AR (Hs00171172_m1), probasin (Mm00444381_m1), NKX3.1 (Mm00440479_m1) and p21 (Hs00355782_m1). A Student's *t*-test was used for statistical analysis of qPCR results, and *P* values less than 0.05 were considered significant.

Histology and immunohistochemistry. The ventral, dorsolateral and anterior prostate lobes were dissected, fixed in 4% formaldehyde in PBS for 24 h, processed and embedded in paraffin. Sections (7 μm) were processed for H&E staining. Prostate hyperplasia is characterized by the proliferation of luminal cells with no cytological atypia, but contain small foci with two or three layers of cells. The PIN lesions were graded using the nomenclature and criteria developed previously[31]. In brief, high-grade PIN is characterized by an intraglandular proliferation of crowding cells with atypia, and cribriform formation or the development of multi-layered solid glandular structures. Invasive adenocarcinoma is characterized by the proliferation of atypical cells that break the basal membrane and invade through the prostatic stroma. The quantitative results of tumour progression are from two random slides of each mouse in a total 10 pairs of the entire animal cohort. For immunohistochemistry, the slides were processed using a citrate-buffer-based antigen retrieval and the avidin–biotin peroxidase immunohistochemistry method. Sections were counterstained with 4′,6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich). Quantification of tumour malignancy was done as follows: for Ki67 staining, positive cells were counted after imaging with a Zeiss Axioplan microscope. Three random slides for each animal were counted. The proliferative rate was determined by counting the proliferating cells (Ki67-positive) in a total of 100 cells, and the statistical significance was determined with a two-tailed Student's *t*-test. For lymph node metastasis, we performed CK8 or AR immunostaining in lymph nodes and recorded the percentage of mice that were positive for CK8 or AR, which we considered as lymph node metastasis. For invasive adenocarcinoma identification, we performed SMAα immunostaining. If the gland or acini displayed a loss or breakdown of the continuous layer of basal membrane (SMAα-positive) and the appearance of AR- or CK8-positive epithelial cells invading into the stromal compartment, we consider it as microinvasive cancer. The statistical significance between groups was determined with a two-tailed Fisher extract test.

Western blot and chromatin immunoprecipitation assay. Total proteins were extracted from cells following standard protocol[11]. Nuclear proteins were extracted using the NE-PER kit (Pierce). Protein concentration was measured using the BCA protein assay kit (Thermo Scientific). The primary antibodies used in this study were as follows: COUP-TFII (Perseus Proteomics), β-actin (Santa Cruz), phophorylated-H3 (Cell Signaling), PCNA (Santa Cruz), phospho-AKT403 (Cell Signaling), pan AKT (Cell Signaling), SMAD4 (Santa Cruz), p21 (BD), phospho-SMAD2 (Cell Signaling), SMAD2 (Zymed), SMAD3 (Zymed), cyclin D1 (Sigma), p15 (Santa Cruz) and p53 (Santa Cruz). Horseradish peroxidase (HRP)-conjugated secondary antibodies were purchased from DAKO. Signals were visualized with the Super Signal West Pico Chemiluminescent Substrate kit (Pierce). Chromatin immunoprecipitation assays were performed as described as previously[11] according to the protocol provided by Millipore. All the results are obtained from three independent samples and statistical significance was determined by a two-sided Student's *t*-test.

Immunoprecipitation and western blot analysis. Immunoprecipitation was performed as described previously[3]. Endogenous or epitope-tagged proteins were immunoprecipitated from cell lysates by the appropriate antibody affinity gel as indicated in the text and figure legends. After extensive washes, immunoprecipitated proteins were eluted in SDS sample loading buffer (Bio-Rad), separated by SDS–PAGE, transferred to nitrocellulose (Pierce) and detected in western blots with appropriate primary antibodies coupled with HRP-conjugated secondary antibody by chemiluminescence (Pierce).

GST fusion protein, *in vitro* protein binding and pull-down assays. GST fusion proteins were prepared using a commercial kit (Amersham Pharmacia Biotech). Target protein was pre-cleared with 5 μg of GST protein for 1 h and then incubated with 2 μg of various GST-fusion proteins for 2 h in the binding buffer (50 mM Tris-HCl, pH 7.5, 120 mM NaCl, 2 mM EDTA, 0.1% NP40). Proteins bound to GST-fusion proteins were retrieved on glutathione–sepharose beads, separated by SDS–PAGE and examined by western blot.

Electrophoretic mobility shift assay. Nuclear proteins from PC3 cells treated with TGF-β for 1 h were extracted using NE-PER kit (Pierce). Complementary oligonucleotides corresponding to the p21 SBE and its mutants were annealed and end labelled with [γ-^{32}P]ATP. The binding reactions were performed and analysed on a 5% non-denaturing gel. For supershift studies, 1 μg of antibody against SMAD4 (Santa Cruz) was preincubated for 20 min on ice, and DNA–protein complexes were visualized by autoradiography.

Microarray analysis. For microarray analysis, total RNA was isolated from control and COUP-TFII-knockdown PC3 cells. The Microarray Core at Baylor

College of Medicine performed the microarray hybridization with Affymetrix Human Genome U133 Plus 2.0. Gene information for all probes was annotated based on 'Human133_2.annot.csv' downloaded from the Affymetrix website. Raw data can be found in the GEO database as accession number GSE33182. The list of significance was operated by setting a false discovery rate (FDR) threshold at level of 0.05. All differentially expressed gene lists generated as described above were further analysed with the Ingenuity Pathways Analysis program (http://www.ingenuity.com/index.html) to identify canonical pathways, and molecular and cellular functions enriched in the related gene lists.

Clinical outcome analysis using GSE1064 data set. We defined a gene expression signature for COUP-TFII activity using an external human tumour data set and determined the correlation of COUP-TFII signature patterns with time to prostate cancer-specific death in patients. The COUP-TFII gene signature was derived from our own gene expression profile data set, of cells with or without COUP-TFII knockdown (GEO accession GSE33182, from ref. 21). The COUP-TFII gene signature (defined using fold change >1.5 for each *COUP-TFII* siRNA sample profile versus each control sample profile) consisted of 1,209 Affy probe sets (representing 829 unique genes) induced with COUP-TFII knockdown and 1,328 probe sets (891 genes) repressed with COUP-TFII knockdown (that is, the expression of 891 genes is induced by COUP-TFII, and 829 genes are repressed by COUP-TFII; Supplementary Data File). We used the previously described '*t*-score' metric[32–36] to define the COUP-TFII gene signature within profiles from an external human tumour data set (GEO accession GSE10645, using the 'human cancer panel', or GPL5858 platform). In brief, the *t*-score was defined for each external profile as the two-sided *t*-statistic comparing, within the profile, the average of the COUP-TFII-induced genes with the average of the COUP-TFII-repressed genes (genes within the GSE10645 data set were first centred to standard deviations from the median). The GSE10645 mRNA profile data set consisted of genes in a focused panel of 488 (which did not include *COUP-TFII* mRNA itself), of which 80 were represented in our COUP-TFII signature; for the 80 genes (47 induced by *COUP-TFII* siRNA and 33 repressed), the *t*-score contrasted the patterns of the COUP-TFII-induced genes against those of the COUP-TFII-repressed genes, to derive a single value denoting coordinate expression of the two gene sets (Supplementary Fig. 7). The univariate Cox proportional hazards analysis, multivariate Cox analysis, and Kaplan–Meier survival analysis were used to determine the correlation of COUP-TFII signature with patient outcome as described below. Furthermore, we performed a univariate analysis for each feature in the GSE10645 data set, and found the overall FDR to be reasonable. Out of 502 total features represented on the focused array, 131 had a univariate Cox $P < 0.01$, yielding an estimated FDR of $(502 \times 0.01)/131 = 3.8\%$. This indicates that there are widespread correlation patterns with outcome represented in the data set, similar to what one might find for other cancer profiling data sets. We also tested the ability of a random gene signature to predict outcome, using simulation testing in which *t*-scores were computed for each of 10,000 randomly generated signatures of the same gene number as the actual signature; for 95.8% of the random signatures, the correlation with worse patient outcome (as assessed by univariate Cox β value) did not exceed that for the actual signature.

Correlation of COUP-TFII signature with PSA recurrence. In addition to the GSE10645 (ref. 21) data set, we examined the signature in three other gene array data sets of human prostate tumours, from refs 37–39. In these other data sets, the outcome endpoint was PSA recurrence, whereas that for the GSE10645 data set was prostate-cancer-specific death. For multiple human tumour gene array data sets including GSE10645, we noted significant overlap between the genes repressed by COUP-TFII knockdown and the genes associated with worse outcome (Supplementary Table 4). However, when computing signature *t*-scores for each of the four data sets, the signature correlated with worse outcome for GSE10645 (ref. 21) (using prostate-cancer-specific death as the outcome measure) and ref. 37 data sets, but not with the refs 38 and 39 data sets, and not with time to PSA recurrence in the ref. 37 data set (Supplementary Table 5). Therefore, we conclude that the signature correlates with shorter time to prostate-cancer-specific death, with time to recurrence event having a weaker association.

Analysis of the enrichment of TGF-β-induced genes in COUP-TFII-depleted PC3 cells. A one-sided Fisher's exact test determined the significance of gene overlap between our COUP-TFII gene signature (as described above) and a previously generated gene signature of TGF-β activity (as reported previously[19]). Both data sets were generated using the Affymetrix U133 Plus 2 platform; therefore, the mapping between the two gene signatures was based on Affymetrix probe set identifier, with the gene population being the entire set ($\sim$54,675) represented on the gene array.

Statistical analysis. All experiments were performed using 3–15 mice or three independent repeated experiments from cells. Data are presented as mean and s.e.m. Spearman correlation coefficients were used to evaluate the relationships between COUP-TFII and clinicopathological variables or gene expression. The predictive value of COUP-TFII expression univariately with other clinical and pathological variables was analysed using the Cox proportional hazards regression model and the hazard ratio, and 95% confidence intervals were computed. The multivariate Cox proportional hazards model with the five-gene signature was used to estimate the coefficients of individual genes, which combined the five-gene expression levels into an integrated risk score model defined. Kaplan–Meier survival curves for different levels of COUP-TFII or COUP-TFII signature were also plotted. All analyses are performed with statistical software SPSS 19.0 (IBM, SPSS Statistics) or WinStat for Excel (R. Fitch software). Statistical significance was determined by a Student's *t*-test, Spearman correlation coefficients test, log-rank test or Fisher's exact test. For all statistical tests, the 0.05 level of confidence (two-sided) was accepted for statistical significance.

26. Agoulnik, I. U. *et al.* Androgens modulate expression of transcription intermediary factor 2, an androgen receptor coactivator whose expression level correlates with early biochemical recurrence in prostate cancer. *Cancer Res.* **66,** 10594–10602 (2006).
27. Wu, S. P., Lee, D. K., Demayo, F. J., Tsai, S. Y. & Tsai, M. J. Generation of ES cells for conditional expression of nuclear receptors and coregulators in vivo. *Mol. Endocrinol.* **24,** 1297–1304 (2010).
28. Wang, S. *et al.* Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. *Cancer Cell* **4,** 209–221 (2003).
29. Feng, X. H., Liang, Y. Y., Liang, M., Zhai, W. & Lin, X. Direct interaction of c-Myc with Smad2 and Smad3 to inhibit TGF-β-mediated induction of the CDK inhibitor $p15^{Ink4B}$. *Mol. Cell* **9,** 133–143 (2002).
30. Li, L. *et al.* The nuclear orphan receptor COUP-TFII plays an essential role in adipogenesis, glucose homeostasis, and energy metabolism. *Cell Metab.* **9,** 77–87 (2009).
31. Park, J. H. *et al.* Prostatic intraepithelial neoplasia in genetically engineered mice. *Am. J. Pathol.* **161,** 727–735 (2002).
32. Creighton, C. J. *et al.* Integrated analyses of microRNAs demonstrate their widespread influence on gene expression in high-grade serous ovarian carcinoma. *PLoS ONE* **7,** e34546 (2012).
33. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2011).
34. Creighton, C. J. *et al.* Insulin-like growth factor-I activates gene transcription programs strongly associated with poor breast cancer prognosis. *J. Clin. Oncol.* **26,** 4078–4085 (2008).
35. Creighton, C. J. *et al.* Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl Acad. Sci. USA* **106,** 13820–13825 (2009).
36. Luo, J. *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137,** 835–848 (2009).
37. Glinsky, G. V., Glinskii, A. B., Stephenson, A. J., Hoffman, R. M. & Gerald, W. L. Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113,** 913–923 (2004).
38. Yu, Y. P. *et al.* Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J. Clin. Oncol.* **22,** 2790–2799 (2004).
39. Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18,** 11–22 (2010).

doi:10.1038/nature11676

Recombination-restarted replication makes inverted chromosome fusions at inverted repeats

Ken'Ichi Mizuno[1], Izumi Miyabe[1], Stephanie A. Schalbetter[1], Antony M. Carr[1] & Johanne M. Murray[1]

Impediments to DNA replication are known to induce gross chromosomal rearrangements (GCRs) and copy-number variations (CNVs). GCRs and CNVs underlie human genomic disorders[1] and are a feature of cancer[2]. During cancer development, environmental factors and oncogene-driven proliferation promote replication stress. Resulting GCRs and CNVs are proposed to contribute to cancer development and therapy resistance[3]. When stress arrests replication, the replisome remains associated with the fork DNA (stalled fork) and is protected by the inter-S-phase checkpoint. Stalled forks efficiently resume when the stress is relieved. However, if the polymerases dissociate from the fork (fork collapse) or the fork structure breaks (broken fork), replication restart can proceed either by homologous recombination or microhomology-primed re-initiation[4,5]. Here we ascertain the consequences of replication with a fork restarted by homologous recombination in fission yeast. We identify a new mechanism of chromosomal rearrangement through the observation that recombination-restarted forks have a considerably high propensity to execute a U-turn at small inverted repeats (up to 1 in 40 replication events). We propose that the error-prone nature of restarted forks contributes to the generation of GCRs and gene amplification in cancer, and to non-recurrent CNVs in genomic disorders.

In eukaryotes, multiple origins are licensed but only a subset fire. If one fork collapses, replication is completed by a converging fork[6]. Alternatively, if both converging forks collapse, dormant origins can fire to rescue the situation[7]. However, when converging forks collapse without an intervening dormant origin, that is, at a fragile site[8,9], or if a single fork collapses at a unidirectionally replicated locus[10], one replisome will likely be rebuilt by HR. To study replication-fork collapse and restart we use a programmed replication-terminator sequence (*RTS1*) to arrest the replisome at a defined genomic locus in fission yeast[11,12]. Fork arrest at *RTS1* is controlled by regulating *rtf1*$^{+}$ transcription[11]. Rtf1, a Myb-like DNA-binding protein, is required for arrest at *RTS1*. After induction, >90% of forks arrest at *RTS1* and require HR proteins to restart[13].

In *Schizosaccharomyces pombe* collapsed forks restart by an HR-dependent, but double-strand-break (DSB)-independent, mechanism[12,13]. Our model (Supplementary Fig. 1) suggests that HR proteins associate with the nascent strand behind the collapsed fork and subsequent strand invasion at the collapse site facilitates accurate HR-dependent fork restart with the correct template. However, if a DNA sequence homologous to the collapse site is nearby, an erroneous strand invasion can occur such that replication reinitiates ectopically. This leads to non-allelic homologous recombination (NAHR)[11–13]. When the homologous sequences are in an inverted-repeat orientation, NAHR associated with inaccurate restart results in acentric and dicentric isochromosomes[13]. We also observed that, when the fork barrier sequence formed the flanking regions of a small palindrome[12], GCR rates increased ~tenfold (contrast the two constructs RuraR and RuiuR, in which a second inverted copy of *ura4*$^{+}$ has been inserted; Supplementary Fig. 2). The main distinction between the two constructs is that RuiuR contains *RTS1* in context of the 5.3-kilobase (kb) palindrome as opposed to an inverted repeat separated by 1.8 kb (RuraR). We thus speculated that, upon NAHR, branch migration of the invading strand (which is not possible in the inverted-repeat construct) formed a single Holliday junction at the palindrome centre, which drove the increased chromosome rearrangement.

To prevent the predicted half-crossover migrating in RuiuR to the palindrome centre we replaced 550 base pairs (bp) of the centromere-proximal *ura4*$^{+}$ gene with 0.2 or 1.8 kb of *his3*$^{+}$, creating Rpal1R and Rpal2R, respectively (Fig. 1a). To prevent any possibility of rearrangement by NAHR we created two further constructs in which the telomere-proximal *RTS1* sequences of Rpal1R and Rpal2R were replaced with three copies of the ribosomal DNA fork barrier sequence, *TER2/3*, to create Tpal1R and Tpal2R. In these constructs, 'T' represents the ribosomal DNA barrier, 'R' represents *RTS1* and 'pal' refers to the presence of the inverted 1.2-kb repeat. *TER2/3* serves simply to pause the converging fork, allowing more time for the *RTS1*-collapsed fork to restart. *TER2/3* differs from *RTS1* in sequence and arrests forks in an Rtf1-independent fashion[14,15] (Supplementary Fig. 3a, b). Unlike *RTS1*, where forks collapse and require HR to rebuild the replisome for restart, forks pause transiently at *TER2/3*, do not require HR for resumption, and the arrest site does not accumulate Rad52. GCRs are thus not induced (Supplementary Fig. 3c–f).

To assay for GCRs, fork arrest was induced at *RTS1* by inducing Rtf1 (*TER2/3* arrest is constitutive) and genomic DNA was analysed by Southern blot. We predicted no GCRs in strains with a single *RTS1* (Tpal1R and Tpal2R) and were interested in establishing whether double-*RTS1* constructs (Rpal1R and Rpal2R) appreciably reduced GCR levels (Fig. 1a). Notably, all four constructs generated GCRs in an *RTS1* fork-arrest-dependent manner (Fig. 1b, c). Double-*RTS1* systems accumulated 25–30% GCRs, similar to that observed in RuiuR, whereas single-*RTS1* strains showed ~5–15% rearrangement.

NAHR between *RTS1* sequences occurring upon restart should produce dicentric chromosomes with an expected BglII fragment of 15.2 kb (Rpal1R) or 15.8 kb (Rpal2R); however, the observed fragment lengths were 14.8 kb and 18 kb, respectively. These correspond to double the size from the centromere-proximal BglII site to the palindrome centre. Identical-size fragments are observed in the corresponding single-*RTS1* constructs. These data suggest a novel mechanism of chromosomal rearrangement, in which the collapsed replication fork resumes accurately with the correct template, but later reverses the orientation of DNA replication (U-turn) as it replicates through the palindrome centre. This leads to isodicentric chromosome formation.

To characterize the effect of palindrome size in promoting restarted-fork-dependent GCRs, a series of constructs was made in which the palindrome size (P(bp)) varied between P(74) and P(2,400), but its centre of symmetry remained a constant distance from the site of fork restart (Fig. 2a). All constructs contained the 14-bp interrupting sequence at the palindrome centre. To establish GCR levels, genomic DNA was analysed by Southern blot using two flanking probes, probe pA or pB.

For P(2,400), pB revealed that ~10% of the DNA corresponded to the rearranged product, migrating at 14.4 kb (Fig. 2b). This is twice 7.2 kb, the distance from the palindrome centre to the centromere-proximal

[1]Genome Damage and Stability Centre, University of Sussex, Brighton, East Sussex BN1 9RQ, UK.

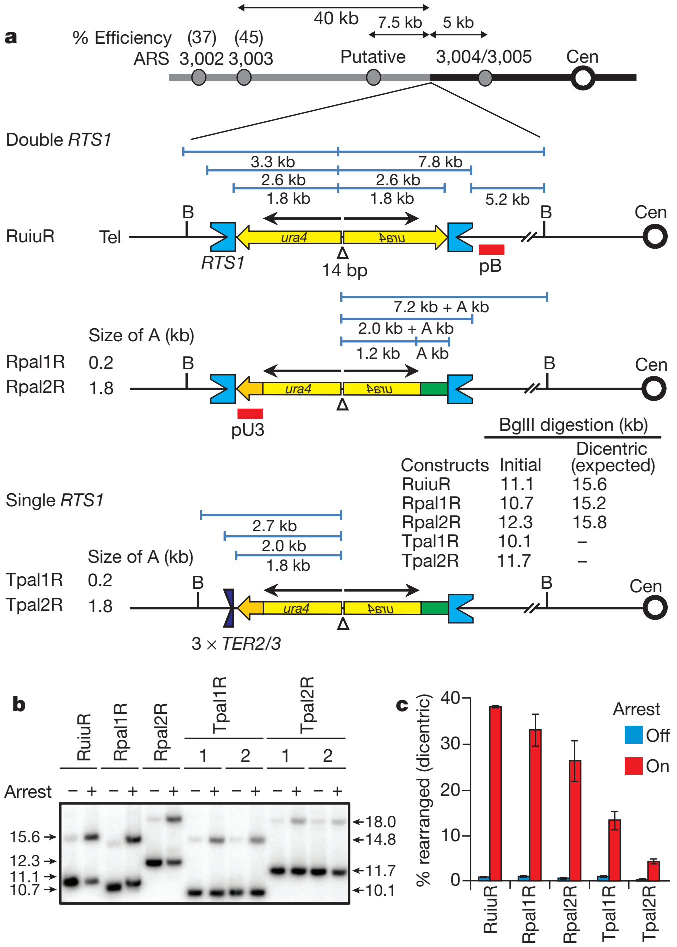


Figure 1 | Alternative mechanism for inverted chromosomal fusion. **a**, Cartoon of double- and single-*RTS1* constructs. Replication origins (ARS) on chromosome 3, their distance from *ura4* and predicted efficiencies are shown. Cen denotes centromere 3. Concave light blue and dark blue boxes represent *RTS1* or 3 × *TER2/3* as indicated. Yellow arrows/boxes represent *ura4* sequences and green box represents *his3* sequences. Black arrow indicates inverted repeats of the *ura4* sequences. Open triangle shows 14-bp interrupting sequence at the palindrome centre. Red bars represent probes. B indicates BglII restriction site. Sizes of initial and predicted dicentric chromosome BglII fragments generated by replication-template exchange are shown for each relevant strain. A kb indicates size of inserted unique sequence (green). **b**, A representative Southern blot of double- and single-*RTS1* strains with arrest off or arrest on. Genomic DNA was digested with BglII and probed with pB. **c**, Quantification of rearranged fragment in **b**. Mean and ± s.d. of values are calculated from at least three independent experiments.

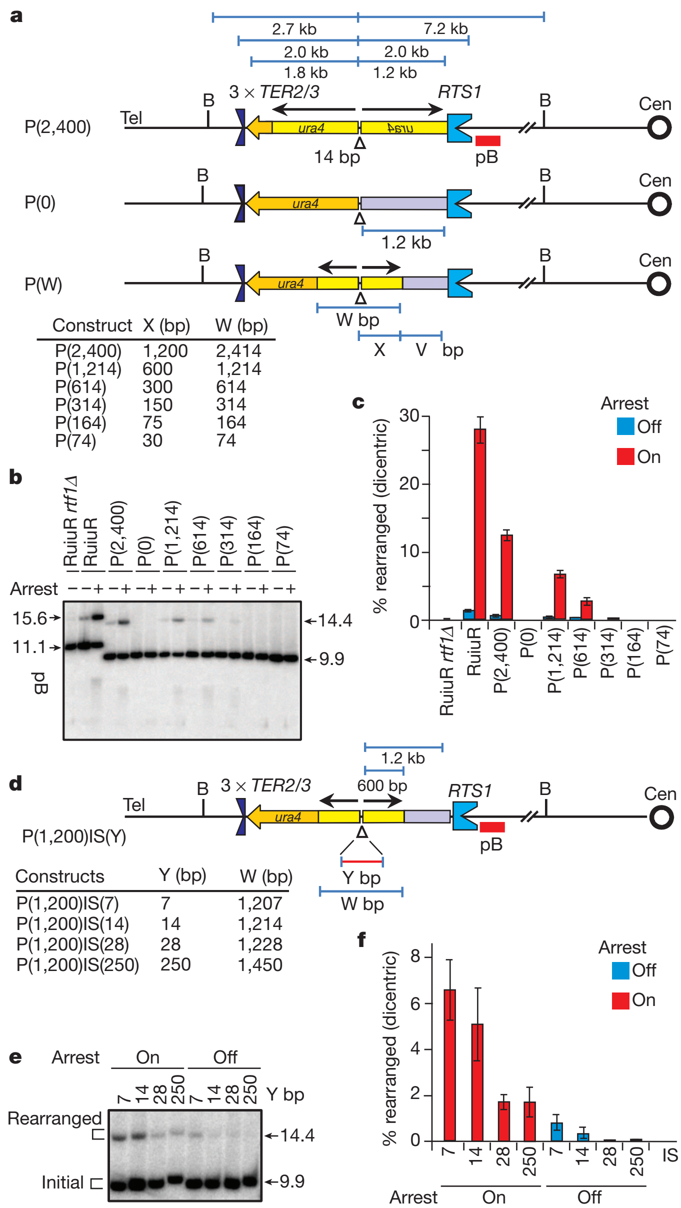


Figure 2 | Rearrangement frequency is dependent on the repeat size and interrupting-sequence size. **a**, Cartoon of constructs with varying repeat size. P(2,400), P(0) (no inverted repeat) and intermediates P(W) are indicated as in Fig. 1a. W represents the size of the whole palindrome in base pairs. X denotes the size of the *ura4* fragment creating the inverted repeat. Grey box indicates heterologous sequence (V). The sum of X and V is always 1,200 bp. **b**, Southern blot analyses of P(W) strains for arrest off or arrest on. *rtf1Δ* indicates strain deleted for the *rtf1* gene. Genomic DNA was digested with BglII and probed with pB. **c**, Quantification of rearranged fragment in **b**. **d**, Cartoon of constructs varying interrupting sequence (IS) (P(1,200)IS(Y)), indicated as in Fig. 2a, in which Y represents the size (in base pairs) of the interrupting sequence. **e**, Southern blot analysis of P(1,200)IS(Y) strains for arrest off or arrest on. Southern blot was performed as described in Fig. 1b. **f**, Quantification of rearranged fragment in **e**, as described in Fig. 1c. Mean and ± s.d. of values in **c** and **f** are calculated from at least three independent experiments.

BglII site. pA revealed a similar proportion of a 5.4-kb fragment (twice 2.7 kb, the distance between the palindrome centre and the telomere-proximal BglII site; Supplementary Fig. 4). These rearranged products correspond, respectively, to dicentric and acentric isochromosomes (data not shown). As expected, P(0), which has no inverted repeats, showed no detectable rearrangement products, confirming that the 14.4-kb and 5.4-kb signals do not represent replication intermediates or broken forks. All induced rearrangements were dependent on replication-fork arrest at *RTS1* and the percentage of GCR was dependent on palindrome size (Fig. 2b, c and Supplementary Fig. 4b, d). P(314) was the smallest palindrome allowing GCR detection by Southern blot analysis.

Palindromes are prone to forming secondary structures, including cruciforms (double-strand DNA) and hairpins (single-strand DNA). Secondary-structure formation is influenced by interrupting-sequence size[16,17]. To establish whether GCR formation was related to interrupting-sequence size, we used P(1,214) as a base construct and varied the interrupting sequence (Fig. 2d). A 7-bp interrupting sequence showed slightly higher GCR levels than a 14-bp interrupting sequence. A 28-bp interrupting sequence reduced levels by ~threefold, to 2%. This did not reduce further when the interrupting sequence was increased to 250 bp (Fig. 2e, f). These results indicate that a potential for structured DNA formation promotes restarted fork U-turn, but is not essential.

The data from Fig. 1b, c suggested that the distance of the palindrome centre from the site of restart influences the U-turn frequency. To clarify this, a further series was constructed in which different sizes

of heterologous sequence separated a 1.2-kb palindrome and *RTS1* (Fig. 3a). The maximum amount of GCR was observed when the palindrome directly abutted *RTS1* (~8%). As the distance between the palindrome and *RTS1* was increased to ~1.5 kb, the GCR level decreased ~fourfold, to 2%. Further extension of the distance did not result in further decreases (Fig. 3b, c). These data indicate that the restarted fork is initially highly error prone, but matures as it travels the first few kilobases. However, the constant rate of GCR observed in constructs separating the palindrome from *RTS1* by >1.5 kb implies that such 'matured' forks are non-canonical and remain error prone.

Although it has not been possible (owing to the limitations of our system) to establish whether recombination-restarted forks become error free over greater distances, it is notable that break-induced replication forks analysed in *Saccharomyces cerevisiae*[18] remain prone to replication slippage[19] and template exchange with homologous chromosomes[20] over tens to hundreds of kilobases. Break-induced replication initiates from a DSB and occurs outside of S phase in G2-arrested cells. Our recombination-restarted forks restart without a DSB intermediate during S phase. It is thus unclear how closely the two systems equate and whether similar replisome configurations underlie restarted replication in both systems.

NAHR occurring during the restart event generates acentric and dicentric isochromosomes containing the intervening sequence originally present between the inverted *RTS1* repeats. Conversely, the isochromosomes generated by recombination-restarted forks executing a U-turn at the palindrome centre contain either two centromere-proximal or two telomere-proximal sequences (defined from the centre of the inverted repeat/palindrome). Thus, we can establish the relative contribution of these two distinct mechanisms in generating isochromosomes from the double-*RTS1* palindrome construct Rpal1R.

Double-*RTS1* constructs (Fig. 1a) potentially cause isochromosome formation by either NAHR or recombination-restarted fork U-turn. The derivatives with a single *RTS1* sequence cannot undergo NAHR and only generate isochromosomes by recombination-restarted fork U-turn. We used Southern blot to distinguish these products (Fig. 4a). The main mechanism for the rearrangement for inverted *RTS1* constructs is recombination-restarted fork U-turn (Fig. 4b), with a minor contribution from NAHR. Therefore, the original palindromic RuiuR construct, which showed a much higher level of isochromosome formation compared with the original RuraR construct[12], does so because of an additional defect associated with forks restarted by HR.

The junction consists of two sister chromatids fused at the repeat centre, suggesting that the recombination-restarted replisome performs a U-turn by exchanging template strands between the repeats. This would be consistent with the nascent strand frequently dissociating transiently from, and then re-annealing to, its template. Synthesis continuing on an incorrect inverted template would result in a 'closed Y' structure at the repeat centre. Intriguingly, large interrupting sequences reduce, but do not eliminate, the rearrangement, suggesting that homology either with or without structured DNA can drive a U-turn.

The genome-rearrangement profiles in cancer are complex, including simple CNVs, chromothripsis (multiple linked rearrangements)[21], translocations and gene amplifications that often initiate from isochromosomes[3]. Similarly, rearrangements in genomic disorders include simple recurrent CNV caused by NAHR during meiosis (or occasionally in mitotic cells), inverted duplication deletions[22], non-recurrent CNVs typified by microhomology (or no homology) at the junction[23], and complex multi-junction events that, in some cases, suggest multiple contiguous replication errors[4,24,25].

Our data show that recombination-restarted forks are error prone, with an unexpectedly high propensity (up to 1 in 40 events) to U-turn between short inverted repeats that can be separated by many kilobases. Likewise, recombination-restarted forks cause increased microhomology-dependent insertions and deletions[26]. Current models for the generation of replication-associated rearrangements almost invariably assume a DSB as the initiating event, which subsequently undergoes an incorrect choice of restart site on the basis of homology or microhomology. We propose that inaccurate replication from forks correctly restarted without a DSB intermediate also makes an important contribution to genome rearrangement. Once a fork is restarted at the correct sequence it is particularly prone to U-turn between inverted repeats. Although our physical assay can detect these events at inverted repeats of ~150 bp, the relationship between repeat size and frequency (Fig. 2c) suggests that shorter repeats will still generate an appreciable rate.

Although we cannot directly establish whether recombination-restarted forks are responsible for genome rearrangements in human cells, our data predicts that isochromosome formation in cancer cells will be elevated at fragile sites, where replication forks are prone to collapse and low-origin density necessitates fork restart. Interestingly, isodicentric chromosome formation and subsequent breakage-fusion-bridge cycle-dependent rearrangements initiate gene amplification[27], and fragile sites have been associated with amplification boundaries as well as other cancer-related GCRs[28]. Similarly, slippage at microhomology by recombination-restarted forks probably contributes to the frequent CNVs associated with cancer[19,26].

Equally, several classes of genomic disorders are compatible with a contribution from recombination-restarted forks: one well-characterized

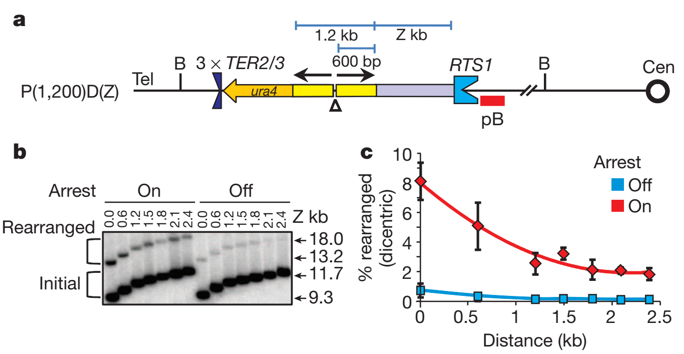


Figure 3 | Fidelity of HR-restarted fork improves with distance. **a**, Cartoon of constructs of varying distance from the restart site (P(1,200)D(Z)) (as indicated in Fig. 2a). The whole palindrome is 1,214 bp. Z indicates distance (D) of the palindrome from *RTS1* in kilobases. **b**, Southern blot analysis of P(1,200)D(Z) strains for arrest off or arrest on. Analysis was performed as described in Fig. 1b. **c**, Quantification of rearranged fragment in **b**, as described in Fig. 1c. Mean and ± s.d. of values are calculated from at least three independent experiments.

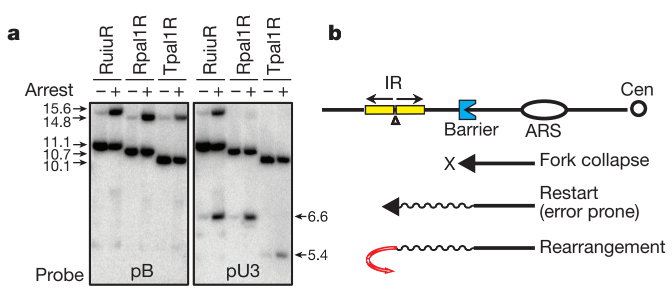


Figure 4 | U-turn at palindrome centre is chief mechanism for inverted fusion in double-*RTS1* constructs. **a**, Southern blot analyses of RuiuR, Rpal1R and Tpal1R for arrest off or arrest on. Genomic DNA was digested with BglII and probed with pB or pU3 (see Fig. 1a). Note that most of the rearrangement in Rpal1R detected by pU3 is acentric 6.6 kb. **b**, Model for error-prone progression of a recombination-restarted replication fork. Oval, blue concave box and yellow box denote replication origin, obstacle and repeat sequences, respectively. When a replication fork collapses, HR restarts the collapsed fork. However, the restarted fork is non-canonical and error prone, causing GCRs at inverted repeats (IR) owing to the execution of a U-turn.

rearrangement involves a triplicated segment embedded within a duplication[29] mediated by small inverted low-copy repeats. 'Duplication-inverted triplication-duplication' involves two 'breakpoint' junctions, one within the repeats and a second showing microhomology. Such a rearrangement can be explained by two distinct events associated with a single-fork restart. Likewise, several other genomic disorders result from inverted duplication deletions[22], which are predicted to be the stabilized events of breakage of an isodicentric chromosome during breakage-fusion-bridge cycles.

METHODS SUMMARY

Standard genetics and molecular biology techniques. Strains were constructed using standard genetic techniques[30]. The *S. pombe* strains used in this study are listed in Supplementary Table 1. Culture conditions, genomic DNA preparation in agarose plugs, Southern blot techniques and the quantification of rearranged DNA were performed as described in refs 12 and 13. Genomic DNA was digested with 150 units of BglII in the recommended buffer. Probes pA and pB are described in ref. 12 as Cen and Tel, respectively. Probe pU3 is a 550-bp fragment of *ura4* genomic DNA digested by EcoRV and SpeI, probe ura is a 1.8-kb *ura4* fragment, and probe ura45 is a HindIII SpeI fragment comprising the $ura4^+$ and $ura5^+$ genes.
***TER2/3* ribosomal DNA fork barrier.** Primers *prr1*-forward (5′-p-AATTCTACTACTATTTTGTGCATTACCCTTACCTTTTTTTTC-3′) and *prr1*-reverse (5′-p-AATTGAAAAAAAAGGTAAGGGTAATGCACAAAATAGTAGTAG-3′) were annealed and ligated. The *TER2/3* consensus sequence is underlined[14]. The ligated DNA was digested with EcoRI and MfeI to eliminate inverted-repeated configuration and size-fractionated by agarose gel, and fragments of ~130 bp (3× tandem repeats) were used to replace the telomere-proximal *RTS1*. The construct was confirmed by sequencing.

Received 22 May; accepted 15 October 2012.
Published online 25 November 2012.

1. Lupski, J. R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14,** 417–422 (1998).
2. Liu, P., Carvalho, C. M., Hastings, P. & Lupski, J. R. Mechanisms for recurrent and complex human genomic rearrangements. *Curr. Opin. Genet. Dev.* **22,** 211–220 (2012).
3. Mondello, C., Smirnova, A. & Giulotto, E. Gene amplification, radiation sensitivity and DNA double-strand breaks. *Mutat. Res.* **704,** 29–37 (2010).
4. Lee, J. A., Carvalho, C. M. & Lupski, J. R. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131,** 1235–1247 (2007).
5. Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* **5,** e1000327 (2009).
6. Petermann, E. & Helleday, T. Pathways of mammalian replication fork restart. *Nature Rev. Mol. Cell Biol.* **11,** 683–687 (2010).
7. Blow, J. J., Ge, X. Q. & Jackson, D. A. How dormant origins promote complete genome replication. *Trends Biochem. Sci.* **36,** 405–414 (2011).
8. Letessier, A. *et al.* Cell-type-specific replication initiation programs set fragility of the *FRA3B* fragile site. *Nature* **470,** 120–123 (2011).
9. Ozeri-Galai, E. *et al.* Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. *Mol. Cell* **43,** 122–131 (2011).
10. Murray, J. M. & Carr, A. M. Smc5/6: a link between DNA repair and unidirectional replication? *Nature Rev. Mol. Cell Biol.* **9,** 177–182 (2008).
11. Lambert, S., Watson, A., Sheedy, D. M., Martin, B. & Carr, A. M. Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell* **121,** 689–702 (2005).
12. Mizuno, K., Lambert, S., Baldacci, G., Murray, J. M. & Carr, A. M. Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism. *Genes Dev.* **23,** 2876–2886 (2009).
13. Lambert, S. *et al.* Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. *Mol. Cell* **39,** 346–359 (2010).
14. Sánchez-Gorostiaga, A., Lopez-Estrano, C., Krimer, D. B., Schvartzman, J. B. & Hernandez, P. Transcription termination factor reb1p causes two replication fork barriers at its cognate sites in fission yeast ribosomal DNA *in vivo*. *Mol. Cell. Biol.* **24,** 398–406 (2004).
15. Krings, G. & Bastia, D. swi1- and swi3-dependent and independent replication fork arrest at the ribosomal DNA of *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **101,** 14085–14090 (2004).
16. Williams, W. L. & Muller, U. R. Effects of palindrome size and sequence on genetic stability in the bacteriophage φX174 genome. *J. Mol. Biol.* **196,** 743–755 (1987).
17. Voineagu, I., Narayanan, V., Lobachev, K. S. & Mirkin, S. M. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc. Natl Acad. Sci. USA* **105,** 9936–9941 (2008).
18. Lydeard, J. R. *et al.* Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. *Genes Dev.* **24,** 1133–1144 (2010).
19. Deem, A. *et al.* Break-induced replication is highly inaccurate. *PLoS Biol.* **9,** e1000594 (2011).
20. Smith, C. E., Llorente, B. & Symington, L. S. Template switching during break-induced replication. *Nature* **447,** 102–105 (2007).
21. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144,** 27–40 (2011).
22. Zuffardi, O., Bonaglia, M., Ciccone, R. & Giorda, R. Inverted duplications deletions: underdiagnosed rearrangements? *Clin. Genet.* **75,** 505–513 (2009).
23. Arlt, M. F., Wilson, T. E. & Glover, T. W. Replication stress and mechanisms of CNV formation. *Curr. Opin. Genet. Dev.* **22,** 204–210 (2012).
24. Liu, P. *et al.* Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell* **146,** 889–903 (2011).
25. Kloosterman, W. P. *et al.* Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. *Cell Rep.* **1,** 648–655 (2012).
26. Iraqui, I. *et al.* Recovery of arrested replication forks by homologous recombination is error-prone. *PLoS Genet.* **8,** e1002976 (2012).
27. Pelliccia, F., Bosco, N. & Rocchi, A. Breakages at common fragile sites set boundaries of amplified regions in two leukemia cell lines K562 - Molecular characterization of FRA2H and localization of a new CFS FRA2S. *Cancer Lett.* **299,** 37–44 (2010).
28. Blumrich, A. *et al.* The *FRA2C* common fragile site maps to the borders of *MYCN* amplicons in neuroblastoma and is associated with gross chromosomal rearrangements in different cancers. *Hum. Mol. Genet.* **20,** 1488–1501 (2011).
29. Carvalho, C. M. *et al.* Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nature Genet.* **43,** 1074–1081 (2011).
30. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194,** 795–823 (1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Hoffman, J. Baxter, M. Neale and members of the Carr and Murray laboratories for discussions. J.M.M. acknowledges Cancer Research UK (CRUK) grant C9601/A9484: A.M.C. acknowledges Medical Research Council (MRC) grant G0600233.

Author Contributions I.M., K.M. and S.A.S. performed experiments. J.M.M., K.M. and A.M.C. wrote the manuscript. All authors contributed to experimental design.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.C. (a.m.carr@sussex.ac.uk).

doi:10.1038/nature11684

Structure of the proton-gated urea channel from the gastric pathogen *Helicobacter pylori*

David Strugatsky[1]*, Reginald McNulty[2]*†, Keith Munson[1]*, Chiung-Kuang Chen[2], S. Michael Soltis[3], George Sachs[1] & Hartmut Luecke[2,4,5,6]

Half the world's population is chronically infected with *Helicobacter pylori*[1], causing gastritis, gastric ulcers and an increased incidence of gastric adenocarcinoma[2]. Its proton-gated inner-membrane urea channel, *Hp*UreI, is essential for survival in the acidic environment of the stomach[3]. The channel is closed at neutral pH and opens at acidic pH to allow the rapid access of urea to cytoplasmic urease[4]. Urease produces NH_3 and CO_2, neutralizing entering protons and thus buffering the periplasm to a pH of roughly 6.1 even in gastric juice at a pH below 2.0. Here we report the structure of *Hp*UreI, revealing six protomers assembled in a hexameric ring surrounding a central bilayer plug of ordered lipids. Each protomer encloses a channel formed by a twisted bundle of six transmembrane helices. The bundle defines a previously unobserved fold comprising a two-helix hairpin motif repeated three times around the central axis of the channel, without the inverted repeat of mammalian-type urea transporters. Both the channel and the protomer interface contain residues conserved in the AmiS/UreI superfamily, suggesting the preservation of channel architecture and oligomeric state in this superfamily. Predominantly aromatic or aliphatic side chains line the entire channel and define two consecutive constriction sites in the middle of the channel. Mutation of Trp 153 in the cytoplasmic constriction site to Ala or Phe decreases the selectivity for urea in comparison with thiourea, suggesting that solute interaction with Trp 153 contributes specificity. The previously unobserved hexameric channel structure described here provides a new model for the permeation of urea and other small amide solutes in prokaryotes and archaea.

Treatment of *H. pylori* infection is becoming less effective as a result of increasing antibiotic resistance, suggesting that a specifically targeted approach to eradicate this organism would be beneficial[5]. Colonization of the acidic mammalian stomach by *H. pylori* depends on the presence of the inner-membrane protein *Hp*UreI, making it a viable clinical target[3]. *Hp*UreI is a proton-gated urea channel that is closed at pH 7.0 and fully open at pH 5.0, enabling the rapid entry of urea into the bacterium[4]. Urease activity (Supplementary Fig. 1) buffers the periplasm to pH 6.1, which is essential for the survival of *H. pylori* at acidic pH[6,7] as a result of the neutralizing capacity of NH_3 and the generation of HCO_3^- by periplasmic α-carbonic anhydrase. Conversely, the closure of *Hp*UreI at neutral pH prevents over-alkalization in the presence of millimolar urea, and consequent cell death[8].

The structure of *Hp*UreI was determined using multiwavelength anomalous dispersion (MAD; Methods). The structure shows an arrangement of six protomers that form a compact hexameric ring (Fig. 1) about 95 Å in diameter and 45 Å in height. The centre of the hexamer is filled with an ordered lipid plug that forms an asymmetric bilayer with electron density for six lipid tails in the periplasmic leaflet and for 18 tails in the cytoplasmic leaflet. This central lipid plug is reminiscent of those reported for other membrane protein oligomers such as bacteriorhodopsin[9]. The main inter-protomer contacts are between TMH1 and TMH2 of one protomer and TMH3 and TMH4 of a neighbouring protomer, a region with conserved residues that are probably important for assembly (Fig. 2a). Native gel electrophoresis

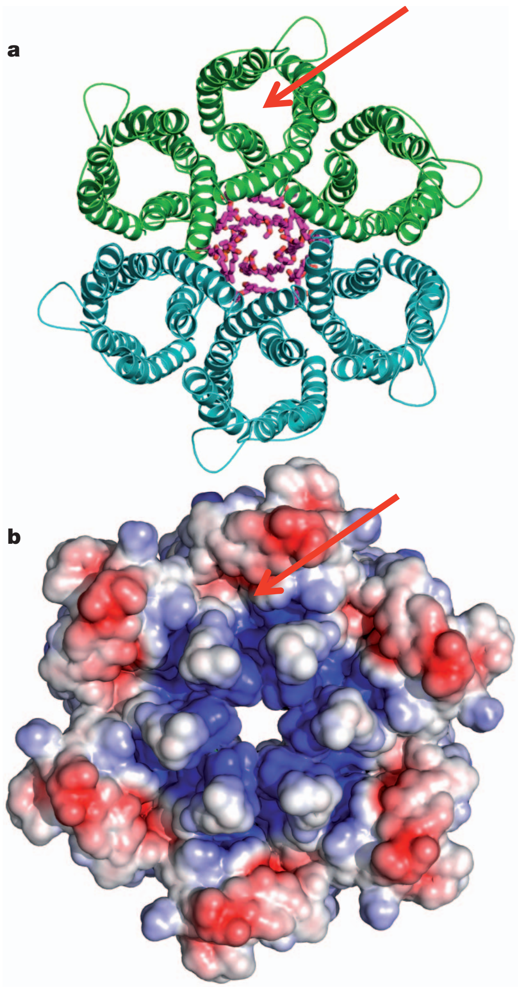


Figure 1 | The *Hp*UreI urea channel hexamer. a, Ribbon diagram of hexamer surrounding the lipids of the central bilayer (purple sticks). The C6 hexamer is generated from the three protomers of one asymmetric unit (green) by the crystallographic two-fold axis (second asymmetric unit in teal). **b**, Electrostatic potential at the periplasmic hexamer surface computed at pH 5.3, the pH at which the crystals were grown (red, $-4kT/e$, blue, $+4kT/e$). The electrostatic potential was calculated with the program APBSmem[21]. The red arrow pinpoints the entrance to one of six urea channels.

[1]David Geffen School of Medicine, University of California Los Angeles and Greater West Los Angeles Health Care System, Los Angeles, California 90073, USA. [2]Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697-3900, USA. [3]Stanford Synchrotron Radiation Lightsource, 2575 Sand Hill Road, Menlo Park, California 94025, USA. [4]Center for Biomembrane Systems, University of California, Irvine, California 92697, USA. [5]Department of Physiology and Biophysics, University of California, Irvine, California 92697, USA. [6]Department of Computer Science, University of California, Irvine, California 92697, USA. †Present address: Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.
*These authors contributed equally to this work.

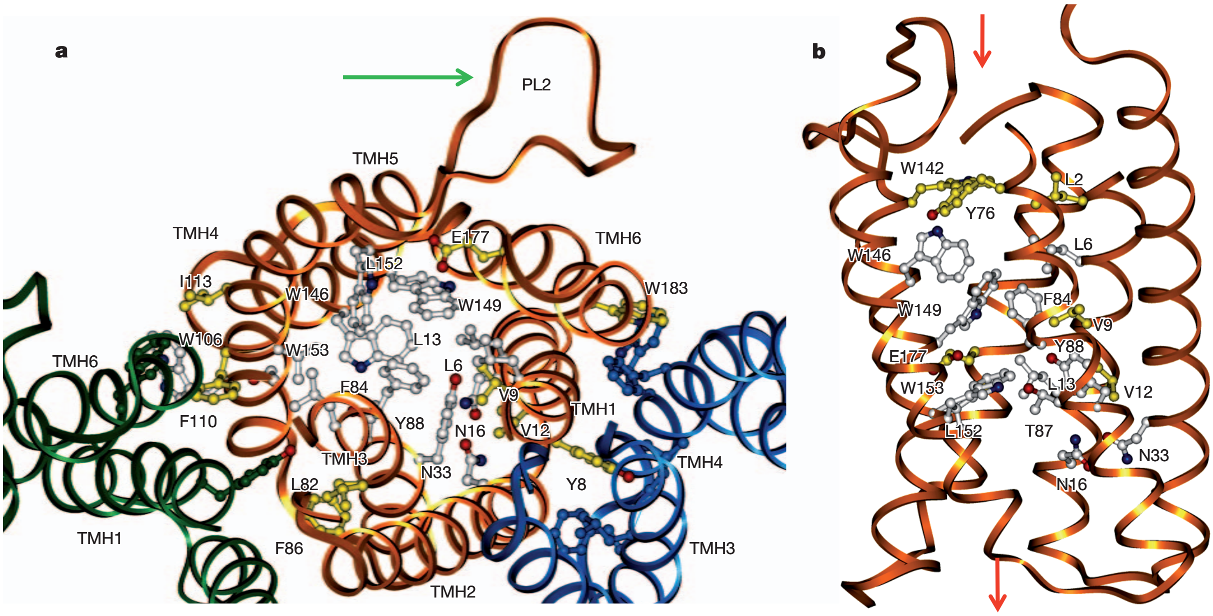


Figure 2 | Residues lining the channel. Residues conserved in the AmiS/UreI superfamily (aligned in Supplementary Fig. 11) are clustered in the channel and at the protomer interface. **a**, View from the periplasm showing the conserved residues of protomer A (gold ribbon) and the open channel of protomers B and C (green and blue ribbons, respectively) when only the protein backbone is shown. For clarity, only selected side chains are labelled. Residues conserved in all members of the AmiS/UreI superfamily are white; additional residues conserved in the subset of known urea channels (including *Hp*UreI) are yellow. The arrow points to the proton-sensing periplasmic loop 2 (PL2) of protomer A. **b**, View parallel to the membrane, with fully conserved residues (white) clustering in the middle of the channel pore, and residues additionally conserved in the urea channels in yellow. The red arrows show the regions of urea entry (top, periplasmic side) and exit (bottom, cytoplasmic side).

of *Hp*UreI confirms that it is a hexamer (Supplementary Fig. 4), in contrast to the previously postulated trimer[10]. An electron microscopy 9-Å projection map of a *Bacillus cereus* amide channel, also a member of the AmiS/UreI family, revealed a similar hexameric arrangement[11]. Thus, the hexamer probably represents the physiological state rather than a crystallization artefact.

Sequence analysis had predicted six transmembrane segments with periplasmic location of both amino and carboxy termini, generating two periplasmic loops (PLs), one between transmembrane helix (TMH)2 and TMH3 (PL1) and another between TMH4 and TMH5 (PL2), as well as a short periplasmic C-terminal segment[4]. The crystal structure shows that each *Hp*UreI protomer is a twisted bundle of six slightly tilted transmembrane helices whose inward-facing side chains define a central channel with a unique arrangement of residues (Fig. 2). Each protomer exhibits non-inverted three-fold pseudosymmetry as demonstrated by the high structural similarity after 120° and 240° rotations around an axis through the centre of the channel (Fig. 3). The repeating motif is a helical hairpin composed of a pair of anti-parallel TMHs connected by a short cytoplasmic loop. However, closer helix–helix packing is observed between helices from adjacent repeated motifs, namely TMH2 and TMH3, TMH4 and TMH5, and TMH1 and TMH6 (Fig. 2a). The mitochondrial ADP/ATP exchanger has a clear internal threefold tandem repeat of a sequence motif in its amino-acid sequence[12], whereas the three-fold pseudosymmetry of *Hp*UreI became apparent only after the crystal structure had been determined. Sequence alignments suggest that this motif is present in eubacteria and archaea (Supplementary Fig. 11), implying an ancient evolutionary origin.

The channel is lined by conserved residues that identify the solute pathway in the AmiS/UreI channel superfamily, and specifically the pathway for urea through *Hp*UreI. The beginning of TMH1 is recessed within the helix bundle and starts with Met 1, whose ordered side chain was crucial for selenomethionine phasing. TMH1 contains 3 of the 11 channel residues that are conserved in the AmiS/UreI superfamily (Fig. 2 and Supplementary Fig. 11), all facing inwards where they define one side of the channel. The longest helix, TMH2, is near the lipid-filled centre, set back from the channel of the hexamer outside the urea pathway, with one conserved residue. This helix starts more than two helix turns earlier than predicted[4]. On reaching the periplasmic side, TMH2 continues for several helical turns until the disordered section of PL1, which contains the engineered His_6 tag. At this point, TMH2 reaches a height above the bilayer that is similar to the height of PL2. TMH3 is situated inwards of TMH2 and is a major contributor to the urea pathway with four conserved residues in the channel, including Phe 84 and Tyr 88, which form part of two constriction sites in the channel (Figs 2 and 4). TMH3 contacts TMH2 of the same protomer as well as the cytoplasmic half of TMH2 from a neighbouring protomer. TMH4, although part of the repeating structural hairpin motif with TMH3, barely contacts TMH3 and instead is in close contact with TMH5. TMH4, PL2, TMH5, cytoplasmic loop 3 and TMH6 constitute the outer, bilayer-facing edge of the *Hp*UreI hexamer. TMH5 contains five tryptophan residues, three of which (Trp 146, Trp 149 and Trp 153) are highly conserved. These line the urea path, with the aromatic side chains of Trp 149 and Trp 153 being the main components of the two constriction sites (Figs 2b and 4). In the middle of the membrane, the side chain of Glu 177 of TMH6 is in a hydrophobic environment and is thus predicted to have an elevated pK_a of 6.7 and to be protonated at pH 5.3, the pH of crystallization. This residue is conserved only in the UreI family of AmiS/UreI members (Supplementary Fig. 12), implying a role in urea transport. It is located to the side of the channel and in a suitable position to hydrogen-bond to urea. Beyond the C-terminal end of TMH6, the last five residues of *Hp*UreI are not helical but bend back towards PL2, where the C-terminal segment is tucked under PL2.

There is probably no high-affinity urea-binding site in the channel, because there is no saturation up to 100 mM urea when expressed in oocytes[4], and the half-saturation concentration in proteoliposomes has been reported to be 163 mM (ref. 10). Although neither structurally nor mechanistically related, *Dv*UT from *Desulfovibrio vulgaris*, a homologue of the mammalian urea transporters, shows half-saturation

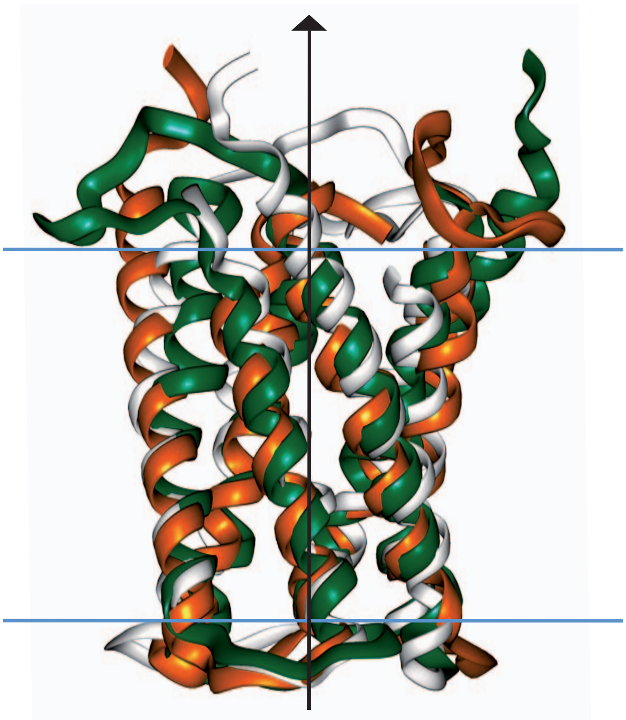


Figure 3 | Structural conservation of a two-helical hairpin motif. The internal three-fold axis of pseudosymmetry is oriented perpendicular to the bilayer (black line) and is the result of a threefold tandem non-inverted repeat of a helical hairpin motif (TMH pairs 1 and 2, 3 and 4, and 5 and 6). In this side view, one *Hp*UreI protomer (gold ribbon) has been overlaid on itself by rotations of 120° (green ribbon) and 240° (white ribbon) around the three-fold axis through the centre of the channel pore. The root mean squared deviations for pairwise overlays of the backbone atoms are 1.52 Å for TMH1 and TMH2 onto TMH3 and TMH4; 1.13 Å for TMH3 and TMH4 onto TMH5 and TMH6; 1.76 Å for TMH5 and TMH6 onto TMH1 and TMH2, and 1.98 Å when all are superimposed simultaneously. Structural conservation is stronger for the cytoplasmic half of the channel. Analysed with the program SymD[22].

concentrations of 2.3 mM urea[13]. However, it is constructed entirely differently, featuring a long (16-Å), narrow hydrophobic channel with inverted two-fold symmetry also found in the mammalian urea transporter UT-B[14]. Both of these contain phenylalanine residues at the entry and exit points of the channel.

*Hp*UreI has much larger periplasmic loops than other members of the AmiS/UreI superfamily, and proton-gated closure of the channel, affected by mutations of histidine residues or carboxylates in the periplasmic domain[15], could be due to loop occlusion of the periplasmic vestibule. A similar effect has been shown for the FocA formate channel[16], but here histidine protonation closes, rather than opens, the channel. For *Hp*UreI at pH 5.3, PL1 is disordered as a result of its engineered His_6 insertion, and PL2 is oriented away from the channel (Fig. 1a), suggesting an open conformation for unimpeded periplasmic entry of urea into the periplasmic vestibule. Urea enters through this irregularly shaped vestibule, which begins roughly at the height of the bilayer edge, defined by a set of largely hydrophobic side chains including Leu 2, Tyr 76, Trp 142 and Trp 146 (Fig. 2b). Next on its way to the cytoplasm, urea must pass through two constrictions, one on each side of Glu 177. Constriction site 1 on the periplasmic side of Glu 177 is defined by the side chains of conserved Leu 6, Phe 84 and Trp 149 (Figs 2b and 4b). Just beyond Glu 177 is constriction site 2, defined by the side chains of conserved Leu 13, Thr 87, Tyr 88, Leu 152 and Trp 153 (Figs 2b and 4b). Trp 153 has its aromatic plane oriented perpendicular to the channel axis, restricting the passage of solutes, and is therefore a candidate for a functional mutation (Supplementary Fig. 8). Its indole side chain can reorient about χ_2 because of minimal steric restraints, enabling the passage of urea and other amides. Similar flexibility of other aromatic side chains in the channel would allow transient π–π stacking of their side-chain surfaces or NH_2–π electrostatic interaction with planar urea to permit selective passage, while at the same time preventing acidification of the cytoplasm by blocking the passage of protons or hydronium. The constrictions in the middle third of the protomer are probably the selectivity filter, which discriminates tenfold between urea and thiourea in oocyte studies[4]. On the cytoplasmic side of constriction site 2, the channel widens again, with Tyr 104 and the hydrogen-bonded pair of Asn 16 and Asn 33 at the beginning of a funnel-shaped vestibule, allowing the exit of urea into the cytoplasm.

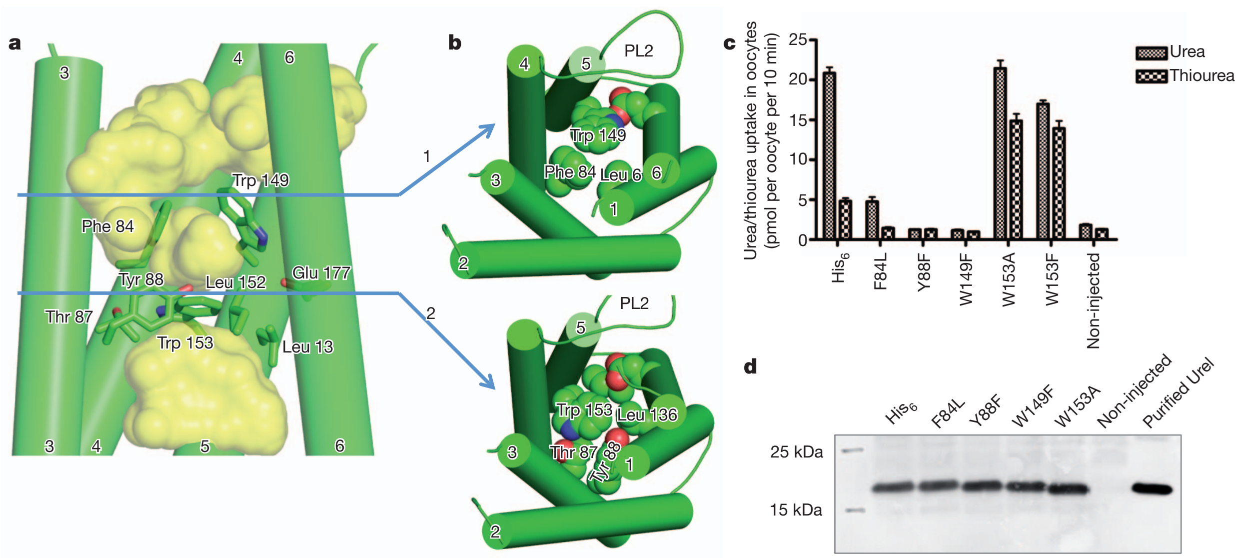


Figure 4 | Views of the *Hp*UreI channel traversed by urea. **a**, Side view with periplasm on top, showing the shape of the channel in yellow and residues conserved in the UreI family (Supplementary Fig. 12). Constriction site 1 is formed largely by the side chains of Leu 6, Phe 84 and Trp 149. On the other side of Glu 177, constriction site 2 is defined by the side chains of Leu 13, Thr 87, Tyr 88, Leu 152 and Trp 153. TMH1 and TMH2 have been removed for clarity. Blue arrows indicate the positions of cross-sections shown in **b**. **b**, Close-ups of constriction sites 1 and 2 viewed from the periplasm. **c**, Channel activity of mutants: mutating Trp 153 to Ala or Phe (constriction site 2) changes selectivity from 4.2:1 (urea:thiourea) for the wild type to 1.5:1 and 1.2:1 for the Ala or Phe mutants, respectively. Mutations at Tyr 88 and Trp 149 inactivate or severely impair the channel, whereas mutation Phe 84⟶Leu retains discrimination ($n = 10$–20; error bars indicate s.e.m.). **d**, Western analysis showing nearly equivalent expression for the different mutants.

Previously, mutations of protonatable residues in the periplasmic domain identified residues important for proton sensing or gating[15]. Here we tested the effect of mutations Phe 84→Leu and Trp 149→Phe in the periplasmic constriction site 1 (Fig. 4b) and Tyr 88→Phe and Trp 153→Ala/Phe in the cytoplasmic constriction site 2 (Fig. 4b) on the discrimination between urea and thiourea (Fig. 4c), a property common to *Hp*UreI, *Dv*UT and UT-B[13,14,16]. The rate of uptake of urea or thiourea was measured in *Xenopus* oocytes. Western analysis confirmed equal expression levels of mutant and wild-type proteins (Fig. 4d). Phe 84→Leu retained the discrimination but with decreased uptake. In the cytoplasmic constriction site, Trp 153→Ala and Trp 153→Phe retained urea transport but lost much of the discrimination between urea and thiourea (Fig. 4c). The selectivity for urea over thiourea is therefore largely determined by Trp 153 in the cytoplasmic constriction site 2 of the channel. The buffering capacity generated by proton activation of *Hp*UreI coupled with urea hydrolysis by the very high levels of urease was estimated on the basis of purified *Hp*UreI (Supplementary Information)[17].

Channel opening also correlates with the acid-dependent recruitment of urease to the cytoplasmic side of the membrane, dependent on *Hp*UreI and essential for survival at pH 3.0 (ref. 18). About two-thirds of the urease is recruited to the plasma membrane on the basis of post-sectioning immunoelectron microscopy at pH 5.0 (ref. 19), confirmed by SDS gel analysis of the membrane fraction of *H. pylori*, which also demonstrates activation of the membrane-recruited urease[20]. Measurement of cytoplasmic pH changes in wild-type and *ureI* deletion mutants also showed that *Hp*UreI increases the bilayer permeability of CO_2 and NH_3/NH_4^+, providing a physiological role for this association by allowing rapid backflow of these molecules to the periplasm[18], thus increasing the rate of periplasmic buffering. These data indicate that a conformational change in the membrane domain of *Hp*UreI may occur with acidification that is transmitted to its cytoplasmic surface.

In contrast with other neutral-solute membrane channels, such as aquaporins[20] and mammalian urea transporters[13,14], which are constructed of two homologous antiparallel halves that generate a long, narrow channel, the parallel three-fold pseudo-symmetric architecture of *Hp*UreI defines a previously unobserved shorter hourglass-shaped channel with a central urea filter and no inverted symmetry. Thus, although both *Hp*UreI and mammalian urea transporters homologous to *Dv*UT selectively allow the passage of urea, the three-dimensional structures and mechanisms of selectivity are distinct.

This first three-dimensional structure of a channel from the AmiS/UreI superfamily reveals a previously unobserved fold with a unique channel architecture able to filter polar organic solutes selectively. As the structure of a validated target for *H. pylori* eradication it may guide the discovery of small-molecule inhibitors, providing the possibility of monotherapy without the use of conventional antibiotics.

METHODS SUMMARY

Crystallization and structure determination. *Hp*UreI was expressed in *Escherichia coli* cells, then purified and crystallized in a mixture of octylglucoside and decylmaltoside, supplemented with *E. coli* polar lipids (Avanti). The structure was solved using SeMet MAD phasing, followed by threefold non-crystallographic averaging, solvent flattening, phase extension, model building and refinement.

Functional assays. Blue native gel electrophoresis indicated a native hexameric structure for *Hp*UreI. The relative rates of transport of urea and thiourea were measured as uptake of tracer in cRNA-injected *Xenopus* oocytes with various mutations.

Full Methods and any associated references are available in the online version of the paper.

Received 24 April; accepted 18 October 2012.

Published online 9 December 2012; corrected online 9 January 2013 (see full-text HTML version for details).

1. Pounder, R. E. & Ng, D. The prevalence of *Helicobacter pylori* infection in different countries. *Aliment. Pharmacol. Ther.* **9,** 33–39 (1995).
2. Peek, R. M. & Blaser, M. J. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nature Rev. Cancer* **2,** 28–37 (2002).
3. Skouloubris, S., Thiberge, J. M., Labigne, A. & De Reuse, H. The *Helicobacter pylori* UreI protein is not involved in urease activity but is essential for bacterial survival *in vivo*. *Infect. Immun.* **66,** 4517–4521 (1998).
4. Weeks, D. L., Eskandari, S., Scott, D. R. & Sachs, G. A. H^+-gated urea channel: the link between *Helicobacter pylori* urease and gastric colonization. *Science* **287,** 482–485 (2000).
5. Graham, D. Y. & Fischbach, L. *Helicobacter pylori* treatment in the era of increasing antibiotic resistance. *Gut* **59,** 1143–1153 (2010).
6. Scott, D. R., Marcus, E. A., Weeks, D. L. & Sachs, G. Mechanisms of acid resistance due to the urease system of *Helicobacter pylori*. *Gastroenterology* **123,** 187–195 (2002).
7. Krulwich, T. A., Sachs, G. & Padan, E. Molecular aspects of bacterial pH sensing and homeostasis. *Nature Rev. Microbiol.* **9,** 330–343 (2011).
8. Meyer-Rosberg, K., Scott, D. R., Rex, D., Melchers, K. & Sachs, G. The effect of environmental pH on the proton motive force of *Helicobacter pylori*. *Gastroenterology* **111,** 886–900 (1996).
9. Luecke, H. *et al.* Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291,** 899–911 (1999).
10. Gray, L. R., Gu, S. X., Quick, M. & Khademi, S. Transport kinetics and selectivity of *Hp*UreI, the urea channel from *Helicobacter pylori*. *Biochemistry* **50,** 8656–8663 (2011).
11. Huysmans, G. H. *et al.* A urea channel from *Bacillus cereus* reveals a novel hexameric structure. *Biochem. J.* **445,** 157–166 (2012).
12. Pebay-Peyroula, E. *et al.* Structure of mitochondrial ATP/ADP carrier in complex with carboxyatractyloside. *Nature* **426,** 39–44 (2003).
13. Levin, E. J., Quick, M. & Zhou, M. Crystal structure of a bacterial homologue of the kidney urea transporter. *Nature* **462,** 757–761 (2009).
14. Levin, E. J. *et al.* Structure and permeation mechanism of a mammalian urea transporter. *Proc. Natl Acad. Sci. USA* **109,** 11194–11199 (2012).
15. Weeks, D. L. & Sachs, G. Sites of pH regulation of the urea channel of *Helicobacter pylori*. *Mol. Microbiol.* **40,** 1249–1259 (2001).
16. Lü, W. *et al.* pH-dependent gating in a FocA formate channel. *Science* **332,** 352–354 (2011).
17. Mobley, H. L., Island, M. D. & Hausinger, R. P. Molecular biology of microbial ureases. *Microbiol. Rev.* **59,** 451–480 (1995).
18. Scott, D. R. *et al.* Cytoplasmic histidine kinase (HP0244)-regulated assembly of urease with UreI, a channel for urea and its metabolites, CO_2, NH_3 and NH_4^+, is necessary for acid survival of *Helicobacter pylori*. *J. Bacteriol.* **192,** 94–103 (2010).
19. Hong, W. *et al.* Medium pH-dependent redistribution of the urease of *Helicobacter pylori*. *J. Med. Microbiol.* **52,** 211–216 (2002).
20. Gonen, T., Sliz, P., Kistler, J., Chen, Y. & Walz, T. Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature* **429,** 193–197 (2004).
21. Callenberg, K. M. *et al.* APBSmem: a graphical interface for electrostatic calculations at the membrane. *PLoS ONE* **5,** e12722 (2010).
22. Kim, C., Basner, J. & Lee, B. Detecting internally symmetric protein structures. *BMC Bioinformatics* **11,** 303–318 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the following for assistance with X-ray data collection at the Stanford Synchrotron Radiation Lightsource (T. Doukov), the Advanced Photon Source (D. Cascio & R. Rajashankar), the Advanced Light Source (J. Nix) and the Swiss Light Source (C. Schulze-Briese). We also thank J. Whitelegge for mass spectrometry characterization of potential heavy atom derivatives; D. R. Scott for the urease assays; B. Hirayama for advice on the oocyte experiments; N. Echols and T. Terwilliger for assistance with the development version of the program Phenix; A. Murzin for fold characterization; T. Silkov for pseudosymmetry analysis; and D. R. Scott, F. Tombola, V. De Souza, K. Luecke, S. Luecke and J. Lanyi for general suggestions. This work was supported by National Institutes of Health (NIH) grants R01AI78000 and P30CA062203, National Cancer Institute institutional training grant 5 T32 CA9054-34, the University of California Irvine Center for Biomembrane Systems (H.L.), NIH grants R01DK53462 and R01DK58333 (G.S.) and the US Veterans Administration (G.S.). Portions of this research were performed at the Stanford Synchrotron Radiation Lightsource, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Stanford University. The Stanford Synchrotron Radiation Lightsource Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the HIN, National Center for Research Resources, Biomedical Technology Program (P41RR001209) and the National Institute of General Medical Sciences.

Author Contributions G.S. and H.L. contributed to the design of the project. D.S. and K.M. expressed, purified and crystallized *Hp*UreI. D.S. conducted the oocyte uptake measurements. C.-K.C. helped with expression and native gels. R.M. and H.L. performed the crystallographic experiments and analysis of data. S.M.S. assisted with aspects of phasing. G.S. and H.L. were responsible for overall project management and wrote the manuscript with K.M.

Author Information The coordinates and structure factors are deposited at the Protein Data Bank under PDB code 3UX4. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.S. (gsachs@ucla.edu) or H.L. (hudel@uci.edu).

METHODS

Engineering a His_6 tag into *Hp*UreI. DNA encoding *Hp*UreI was isolated by PCR from *Helicobacter pylori* strain J99. A His_6 tag was introduced into the protein at various locations to facilitate purification. The engineered proteins with a His_6 tag at the N terminus, in the first periplasmic loop (PL1), the second periplasmic loop (PL2) or at the C terminus were expressed in *Xenopus* oocytes and tested for channel activity. Wild-type *Hp*UreI showed urea uptake of 29.04 ± 1.35 pmol per oocyte per 10 min (mean ± s.e.m.; $n = 9$) and the PL1 His_6 tag (after Gly 61) showed an uptake of 22.36 ± 0.60 pmol per oocyte per 10 min (mean ± s.e.m.; $n = 8$). Mutants with the His_6 insertion at the N terminus, in PL2 or at the C terminus, were inactive. *Hp*UreI with the His_6 tag inserted in PL1 (HpUreI6HisPL1) was subsequently used for expression, purification and crystallization.

***Hp*UreI expression and membrane isolation.** *Small-scale expression.* pET101-UreI6HisPL1 was transformed into *E. coli* C43 (Avidis S.A.). For small-scale expression and crude membrane isolation, bacterial cultures were grown to $D_{600} = 0.8$ and then induced by the addition of 1 mM isopropyl β-D-thiogalactoside (IPTG). After 3 h of induction, cells were harvested by centrifugation at 1,500*g* for 10 min. The pellet was resuspended in a solution of 50 mM Na_2HPO_4 pH 7.4, 1 mM EDTA, 30 μg ml^{-1} DNase I and sonicated to lyse the cells. Cell debris was removed by centrifugation (10,000*g*, 10 min) and membranes were collected (100,000*g*, 45 min) and resuspended in the same buffer without DNase I (40–50 μl). Samples of membrane protein (25 μg, determined by bicinchoninic acid assay (Pierce Biotech, Inc.)) were dissolved in gel loading buffer containing 1% 2-mercaptoethanol and run without boiling on 4–12% SDS–polyacrylamide gels. After transfer to nitrocellulose, western blot analysis was performed with either an anti-UreI or an anti-His-tag antibody (GE).

Large-scale membrane expression for crystallization trials. A bioreactor (BioFlo 110; New Brunswick) containing 10 l Luria–Bertani broth supplemented with 50 mM K_2HPO_4 pH 7.8 and 1.5% (w/v) glycerol was inoculated with 0.2 l of overnight culture of the *E. coli* harbouring a pET101HpUreI6HisPL1 plasmid. When D_{600} reached 0.8–1.0 (about 2 h), *Hp*UreI expression was induced with 1 mM IPTG. Growth was maintained at 37 °C, with a continuous air supply at 5 l min^{-1} and stirring at 500 r.p.m. for 5 h until the cell density reached a D_{600} of 5–7. Cells were then collected by centrifugation and resuspended at 4 °C in 400 ml of buffer containing 50 mM Na_2HPO_4 pH 7.4, 30 μg ml^{-1} DNase and 10 mM 2-mercaptoethanol. All subsequent steps were conducted at 4 °C. The bacterial suspension was passed twice through a French press cell at 20,000 lb in^{-2}. The cell debris was removed (10,000*g*, 20 min) and the remaining membranes were collected (100,000*g*, 2 h) to yield about 4 g of total membrane protein. The membrane pellet was resuspended in 350 ml of storage buffer (10 mM imidazole, 150 mM NaCl, 50 mM Na_2HPO_4 pH 7.4, 65 g l^{-1} glycerol and 2 mM mercaptoethanol) to yield 10–12 mg ml^{-1} membrane protein, which was then flash-frozen in liquid nitrogen and stored at −80 °C.

Labelling with selenomethionine. The pET101UreI6HisPL1 construct was transformed into the methionine auxotroph strain BL21-CodonPlus-RP-X strain (Stratagene) for optimal labelling of *Hp*UreI with selenomethionine in minimal medium. A 10-l culture of M9 minimal medium (60 g l^{-1} Na_2HPO_4, 30 g l^{-1} KH_2PO_4, 10 g l^{-1} NH_4Cl, 5 g l^{-1} NaCl, 0.4% (w/v) glucose, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 0.1 l MEM vitamin solution (Mediatech)) supplemented with 60 μg ml^{-1} selenomethionine, was inoculated with 200 ml of overnight culture grown in the same minimal medium with 0.1 mg ml^{-1} methionine in place of selenomethionine. Protein expression was induced with 1 mM IPTG at $D_{600} \approx 0.8$, and growth was continued for 40 h at 20 °C.

***Hp*UreI purification.** *Solubilization and metal affinity chromatography.* Membranes in storage buffer were collected at 100,000*g* for 90 min, resuspended at 10 mg ml^{-1} in a buffer of 10 mM imidazole, 150 mM NaCl, 50 mM sodium phosphate pH 7.4, and partly dissolved in 2% decylmaltoside (DM; Anatrace) by adding a 20% detergent solution drop by drop and stirring on ice for 30 min. The insoluble material was pelleted (160,000*g*, 30 min) and the supernate was gently agitated overnight with 5 ml of cobalt–agarose resin (TALON resin; Clontech) at 4 °C to bind the His_6-UreI. All the solutions used in subsequent rinsing and elution of the resin contained 150 mM NaCl, 50 mM sodium phosphate pH 7.4, 0.2% DM and 0.1 mg ml^{-1} *E. coli* polar lipids (Avanti) with increasing concentrations of imidazole. The resin was rinsed with ten volumes of buffer with 10 mM imidazole, then 30 volumes with 20 mM imidazole. *Hp*UreI was then eluted with 30 ml of the same buffer with 250 mM imidazole. The eluted *Hp*UreI was concentrated to about 10 mg ml^{-1} by using Amicon 50-kDa filters before gel filtration.

Size exclusion. *Hp*UreI was purified on a Superose 12 column (10/300; GE) with buffer containing 10 mM MES pH 6.5, 150 mM NaCl, 0.2% DM and 0.1 mg ml^{-1} *E. coli* polar lipid extract. The peak fractions were pooled and concentrated to 10 mg ml^{-1} (50-kDa filters; Amicon) for use in crystallization trials.

***Hp*UreI crystallization.** *Hp*UreI protein at 10 mg ml^{-1} in Superose 12 column buffer was diluted to give a solution comprising 1.57 mg ml^{-1} *Hp*UreI protein, 40 mM NaCl, 1 mM TlCl, 10 mM $CaCl_2$, 7% PEG 400, 0.05% decylmaltoside, 2.25% octylglucoside, 0.8 mg ml^{-1} *E. coli* polar lipids (Avanti) and 35 mM MES pH 5.3. This mixture (3.5 μl) was used for hanging-drop diffusion over a reservoir (0.5 ml) of 20% PEG 400 in 0.1 M MES pH 5.3. Crystals grew in 3–4 months at 11 °C and were dehydrated by raising the PEG 400 concentration in the reservoir in increments of 3% at 2-day intervals until the final PEG 400 concentration in the well solution was 33%. The crystal used for collection of the native data set is shown in Supplementary Fig. 1.

X-ray data collection and data reduction. Single crystals were mounted in nylon loops and flash-cooled in liquid nitrogen. X-ray diffraction data collection was conducted at 100 K by collecting 180 diffraction images 1° in width. Data were integrated, scaled and merged with the program XDS[23] (Supplementary Table 1). Because of the large variability of the *c*-axis length (135–152 Å), data could not be merged between multiple crystals, making experimental phasing more challenging. After diffraction data from hundreds of heavy-metal-soaked crystals had been collected without yielding interpretable maps, we resorted to selenomethionine (SeMet) phasing, despite the low abundance of methionine in *Hp*UreI J99 (three residues out of 195, counting the N-terminal methionine). Although fluorescence scans of the initially very small SeMet crystals indicated significant incorporation of selenium, selenium sites could not be located in more than 50 MAD and single-wavelength anomalous dispersion data sets collected with 30° wedges.

The successful MAD data set was collected from a rare larger crystal that took more than 5 months to grow. Three-wavelength MAD diffraction data were collected in 10° wedges at beamline 12-2 at SSRL (Supplementary Table 1).

SeMet MAD phasing. The program SHELXC[25] was used to determine the approximate resolution cutoff for the anomalous signal (when the correlation coefficient between the anomalous differences at the peak and remote wavelengths decreased to below 30%). SeMet sites were initially obtained using the program SHELXD[24]. Of the nine possible sites (three protomers in the asymmetric unit with three SeMet sites per protomer, including all three N-terminal SeMet residues), eight sites were located. From the symmetry of these eight sites, a ninth site could be located, leading to the localization of all selenium anomalous scatterers. The boundary of the *Hp*UreI hexamer and individual helices were readily apparent in the resulting low-resolution maps; however, unsurprisingly, detail of the maps and phasing power was poor. To improve the phases, the program autoSHARP[25] was used. The figure of merit from autoSHARP phasing was used to determine the high-resolution cutoff for the experimental phases. A cutoff of 0.3 for the figure of merit of acentric reflections was used for this estimation, suggesting useful phases to about 4.7 Å resolution (Supplementary Table 2).

Phase improvement and extension. To improve and extend the initial experimental MAD phases, solvent flattening, three-fold non-crystallographic symmetry (NCS) and multi-crystal averaging of data from crystals with different *c*-axis lengths were performed. To obtain initial matrices for NCS averaging, 18 ideal helices were modelled into the experimental electron density representing the three molecules in the asymmetric unit. Subsequently the NCS matrices from molecule A to molecule B and from molecule A to molecule C were obtained by the SSM superpose function of the program COOT[26]. A mask was placed around the electron density of molecule A and threefold NCS averaging and solvent flattening (68.9% solvent) were performed with the program DM[27], while NCS matrices were refined. Phases were extended in an iterative fashion from 4.5 Å to 3.5 Å by using small resolution increments with the programs DM and DMMulti[27]. At this point the right-handed twist of the helices and density for some of the larger side chains could be seen.

Model building and refinement. Density representing each of the six transmembrane helices was cut out with the program PHENIX[28]. An ideal helix was built into this density, slightly curved if required, using the sequence representing each helix. Helix 6 contains a π-bulge and was built by hand with the program COOT. Cytoplasmic loops 1 and 2 (CL1 and CL2) and periplasmic loop 2 (PL2) were initially modelled into density with the program RAPPER[29]. In places where the correct sequence could not be built initially, alanine was used temporarily. To generate the other two molecules of the asymmetric unit, the previously identified NCS matrices were used, followed by rigid-body refinement with the program PHENIX. The model was refined iteratively by cycles of manual adjustments with the program COOT and refinement with the program PHENIX, using TLS refinement with NCS and secondary structure restraints.

On analysing the data with the University of California Los Angeles anisotropy server[30], the data were found to be severely anisotropic with an effective resolution of 3.1 Å in the *a** and *b** directions, but only 3.5 Å in the *c** direction. The data were ellipsoidally truncated and rescaled to minimize the inclusion of poor diffraction data. The model was refined with the newly truncated data by using jelly body refinement with the program REFMAC[31], leading to significantly improved electron density maps that allowed further model improvement. Because of

disorder, no model was built for the majority of periplasmic loop 1 (PL1, residues 59–73), which contains the engineered His_6 insertion.

A final round of refinement was performed with the program REFMAC against the non-truncated data with two TLS groups in each protomer (residues 1–146 and 147–195), tight NCS restraints and a jelly body value of 0.01. The final structure was refined to 3.26 Å with few Ramachandran outliers, small deviations from ideal geometry and predominantly preferred side-chain rotamers (Supplementary Table 1).

SeMet anomalous difference maps. To validate the sequence assignment of the *Hp*UreI crystal structure, additional methionine residues were engineered into various sites in the *Hp*UreI sequence and labelled with SeMet. These proteins were crystallized and their anomalous differences Fourier maps were inspected. For all five engineered sites, the location of the anomalous difference peak was less than 1 Å from the site of the methionine sulphur atom of the final model. The sites were Ile 14, Ala 148, Thr 155, Leu 173 and Ile 191, in addition to the endogenous sites Met 1, Met 14 and Met 127, the last of these being situated in PL2.

Estimation of the number of *Hp*UreI molecules per bacterial cell by western blotting. Bacterial proteins were size-fractionated by SDS–tricine polyacrylamide gel electrophoresis and electroblotted onto a nitrocellulose membrane, followed by immunodetection by enhanced chemiluminescence (Amersham). The anti-*Hp*UreI antibody used for detection and quantification was generated in rabbit against PL1 between TMH2 and TMH3 (CEGAEDIAQVSHHLTSFYGPATG)[32]. Immunoblots were digitized (ScanMaker i800; Microtek) at a resolution of 600 d.p.i. The scanned images were analysed with Kodak 1D software and the amount of *Hp*UreI was quantified with purified *Hp*UreI as the reference. There are 5.3 ng of *Hp*UreI per μg total protein (Supplementary Fig. 2) in 5.4×10^6 cells (measured by colony-forming units). The molecular mass of *Hp*UreI is 21.7 kDa. There are therefore 0.24 pmol of *Hp*UreI per 5.4×10^6 cells; 0.24 pmol of *Hp*UreI $\times 6.0 \times 10^{23}/(5.4 \times 10^6$ cells) is equivalent to about 27,000 channels per cell.

Urease activity assay. Bacteria grown on trypticase soy agar (TSA) plates were suspended in 1 ml of 1 mM phosphate buffer to a final D_{600} of 0.01. Urease activity was measured radiometrically. Bacteria were added to 100 mM sodium phosphate buffer containing 5 mM KCl, 138 mM NaCl, 0.5 mM $MgCl_2$, 1 mM $CaCl_2$, 10 mM glucose and 5 mM [^{14}C]urea with a specific radioactivity of 10 μCi μmol^{-1}. To measure activity in the closed and open states, the pH of the buffer was set to pH 4.5 and 7.4 by mixing various amounts of 100 mM monobasic sodium phosphate and 100 mM dibasic sodium phosphate to the desired pH. The pH of the buffer during the course of the experiment did not change by more than 0.1 pH units. Plastic wells containing 500 mM KOH-soaked filter paper hung from rubber stoppers were used to collect the total $^{14}CO_2$ from the hydrolysis of urea. Urease activity was measured for 30 min at 37 °C with constant agitation. The reaction was terminated with 2.5 M H_2SO_4 to release all ^{14}C as CO_2 and incubated for a further 30 min at 37 °C. The wells were placed in scintillation cocktail (HiIonicFluor; Packard Instruments), and the radioactivity was measured by scintillation counting (1216 RackBeta; LKB Instruments). Protein concentration was determined by the bicinchoninic acid method (Pierce). Urease activity was about 2.3 μmol min^{-1} mg^{-1}, sufficient to hydrolyse all entering urea[32].

Urea flux estimation. To estimate the urea flux per channel per second, the total urease activity of *H. pylori* cells was measured under *Hp*UreI-mediated flux conditions. On the assumption that all urea entering the cell is hydrolysed by the large amount of cytoplasmic urease (about 8–10% of total protein) and knowing the number of *Hp*UreI molecules per cell (27,000; see above), the lower bound of the flux was estimated to be about 200 urea molecules per *Hp*UreI channel per second at 5 mM medium urea, calculated from the rate of urea hydrolysis and the number of channels per organism.

This is a tenfold higher influx of urea than that due to unfacilitated diffusion across phospholipid bilayers (4×10^{-6} s^{-1}) (ref. 33) and it agrees with the tenfold increase in urease activity observed when *Hp*UreI is open. This calculation does not take into account the roughly fourfold increase in urease activity when incubated at acidic pH (ref. 6); hence the rate could be as low as 1,000 s^{-1}.

NH_3 production due to urea influx through *Hp*UreI. Using the calculated flux rate and the number of channels per cell from above it was estimated that at least 5.4×10^6 urea molecules are transported through *Hp*UreI per cell per second at gastric urea concentrations. At pH 5.0, with *Hp*UreI fully open, this can generate about 30 mM NH_3 per second neutralizing capacity in a bacterium with a volume of 0.6 fl.

Blue native gel electrophoresis of *Hp*UreI. *E. coli* membranes were treated with 2% n-dodecyl-β-D-maltopyranoside (DDM) and the soluble fraction was bound to TALON resin. After washes with buffer containing 20 mM imidazole, *Hp*UreI was eluted with a buffer containing 150 mM imidazole, 50 mM sodium phosphate pH 7.4, 150 mM NaCl and 0.01% DDM. The eluate was concentrated on Amicon filters with 10-kDa cutoff to about 5 mg ml^{-1}. Protein sample (5 μl) was mixed with 2 × sample buffer (100 mM Tricine, 30 mM Bis-Tris pH 7.0, 30% glycerol, 0.02% DDM). The cathode buffer was 50 mM Tricine, 15 mM Bis-Tris pH 7.0, and the anode buffer was 50 mM Bis-Tris pH 7.0. The electrophoretic shift of proteins was achieved by including Coomassie blue G250 dye in the cathode buffer. The running conditions were as follows: initial concentration of Coomassie blue G250 was 0.02% in the cathode buffer and the gel was run for 30 min at 100 V and for a further 2 h at 200 V. The concentration of dye was then decreased to 0.002% and the gel was run for a further 2 h at 300 V. High-molecular-mass markers (catalogue number 17-0445-01; Amersham), BSA and trypsin inhibitor (Sigma) were mixed with the same sample buffer (Supplementary Fig. 3).

Solute uptake experiments in *Xenopus laevis* oocytes. Genes encoding wild-type *Hp*UreI (*H. pylori* strain 43504) and His_6-tagged *Hp*UreI (*H. pylori* strain J99 with the His_6 tag in PL1) were cloned into the pcDNA3.1 (Invitrogen) and pET101 (Invitrogen) vectors, respectively. For the experiments on site-specific mutants, substitutions were introduced into the gene encoding His_6-tagged *Hp*UreI with the use of the QuickChange method (Stratagene). Capped and poly(A)-tailed RNA (cRNA) was prepared with the mMessage mMachine T7 Ultra Kit (Ambion). cRNA (50 nl, 1 μg μl^{-1}) was injected into each oocyte (Ecocyte BioScience). Oocytes were incubated for 2 days at 18 °C in Barth's solution (88 mM NaCl, 2.4 mM $NaHCO_3$, 1 mM KCl, 0.4 mM $CaCl_2$, 0.3 mM $Ca(NO_3)_2$, 0.8 mM $MgSO_4$, 10 mM HEPES-Tris pH 7.5). Oocytes were transferred to new vials containing 0.5 ml of Ringer's solution (100 mM NaCl, 2 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$) buffered by 20 mM MES at pH 5.0 for comparison between urea and thiourea uptake through the open *Hp*UreI channel. The transport reaction was started by the addition of 100 μM [^{14}C]urea or 100 μM [^{14}C]thiourea. After 10 min of incubation at 20 °C the solution was aspirated to terminate transport and the oocytes were washed twice with ten volumes of ice-cold radioisotope-free Barth's solution. Oocytes were transferred to individual scintillation vials and dissolved in 10% SDS before the addition of scintillation cocktail.

Western analysis of mutant expression levels. Ten oocytes were solubilized by pipetting into buffer containing 0.5 ml of 0.1 M NaCl, 0.02 M Tris-HCl pH 7.6 and 1% Triton X-100. The yolk was spun down by centrifugation at 20,000*g* for 5 min and the supernatant was aspirated with care, to minimize contamination with floating lipid. Supernatant (20 μl) was size-fractionated by SDS–tricine polyacrylamide gel electrophoresis and electroblotted onto a nitrocellulose membrane followed by immunodetection with alkaline phosphatase (Promega). The anti-*Hp*UreI antibody used for detection was generated in rabbit against PL1 between TMH2 and TMH3 (CEGAEDIAQVSHHLTSFYGPATG). All mutants were expressed at similar levels (Fig. 4d).

23. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 (2010).
24. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. D Biol. Crystallogr.* **64,** 112–122 (2008).
25. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364,** 215–230 (2007).
26. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
27. Cowtan, K. Error estimation and bias correction in phase-improvement calculations. *Acta Crystallogr. D Biol. Crystallogr.* **55,** 1555–1567 (1999).
28. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
29. Furnham, N. *et al.* Knowledge-based real-space explorations for low-resolution structure determination. *Structure* **14,** 1313–1320 (2006).
30. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103,** 8060–8065 (2006).
31. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374,** 300–321 (2003).
32. Scott, D. R. *et al.* Expression of the *Helicobacter pylori ureI* gene is required for acidic pH activation of cytoplasmic urease. *Infect. Immun.* **68,** 470–477 (2000).
33. Orbach, E. & Finkelstein, A. The nonelectrolyte permeability of planar lipid bilayer membranes. *J. Gen. Physiol.* **75,** 427–436 (1980).

doi:10.1038/nature11685

Drosophila NOMPC is a mechanotransduction channel subunit for gentle-touch sensation

Zhiqiang Yan[1]*, Wei Zhang[1]*, Ye He[1], David Gorczyca[1], Yang Xiang[1], Li E. Cheng[1], Shan Meltzer[1], Lily Yeh Jan[1] & Yuh Nung Jan[1]

Touch sensation is essential for behaviours ranging from environmental exploration to social interaction; however, the underlying mechanisms are largely unknown[1]. In *Drosophila* larvae, two types of sensory neurons, class III and class IV dendritic arborization neurons, tile the body wall. The mechanotransduction channel PIEZO in class IV neurons is essential for sensing noxious mechanical stimuli but is not involved in gentle touch[2]. On the basis of electrophysiological-recording, calcium-imaging and behavioural studies, here we report that class III dendritic arborization neurons are touch sensitive and contribute to gentle-touch sensation. We further identify NOMPC (No mechanoreceptor potential C), a member of the transient receptor potential (TRP) family of ion channels, as a mechanotransduction channel for gentle touch. NOMPC is highly expressed in class III neurons and is required for their mechanotransduction. Moreover, ectopic NOMPC expression confers touch sensitivity to the normally touch-insensitive class IV neurons. In addition to the critical role of NOMPC in eliciting gentle-touch-mediated behavioural responses, expression of this protein in the *Drosophila* S2 cell line also gives rise to mechanosensitive channels in which ion selectivity can be altered by NOMPC mutation, indicating that NOMPC is a pore-forming subunit of a mechanotransduction channel. Our study establishes NOMPC as a bona fide mechanotransduction channel that satisfies all four criteria proposed for a channel to qualify as a transducer of mechanical stimuli[3] and mediates gentle-touch sensation. Our study also suggests that different mechanosensitive channels may be used to sense gentle touch versus noxious mechanical stimuli.

Similar to vertebrates, *Drosophila* displays mechanosensation such as gravity sensing, hearing, proprioception, mechanical nociception and gentle-touch sensation[4–11]. To identify sensory neurons for gentle touch we used a previously established behavioural assay[12]. In brief, one side of the larval thoracic segments was gently touched with an eyelash and the behavioural responses were scored (Supplementary Fig. 1 and Supplementary Methods). By expressing the tetanus toxin light chain (TNT) in different classes of dendritic arborization neurons to interfere with synaptic transmission, we found that the behavioural response to gentle touch was reduced by TNT expression in class III (Fig. 1a, b) but not class IV dendritic arborization neurons (Supplementary Fig. 2). In all of the behavioural studies we used a single copy of upstream activation sequence (UAS)-TNT and a single copy of the yeast Gal4 driver, which may not have been sufficient to completely eliminate the synaptic output of class III dendritic arborization neurons. Alternatively, the residual larval behavioural response to gentle touch may indicate the involvement of additional sensory neurons.

As dendrites of five class III dendritic arborization neurons tile 70–80% of each abdominal hemisegment[13] (Supplementary Fig. 3), a *Drosophila* larva could use these neurons to detect gentle touch on most of its body. Extracellular recordings of class III neurons in abdominal segments of fillet preparations (Supplementary Fig. 4) revealed that a single touch causing a displacement of the body wall by as little as 10 μm—about one-hundredth of the width of a third-instar larva—induced a burst of action potentials (Fig. 1c). Progressively larger displacements elicited greater numbers of action potentials in a graded fashion (Fig. 1d, i). In addition, a single touch causing a 40-μm displacement induced a large calcium response in class III neurons (Fig. 1e). Similar to the five class III dendritic arborization neurons in each abdominal hemisegment, class III neurons in thoracic segments showed touch responses (Supplementary Fig. 5). By contrast, the same touch stimulation had no effect on class IV dendritic arborization neurons that are mechanical nociceptive neurons[2,14] (Supplementary Fig. 6).

Class III neurons adapted quickly to a prolonged stimulus with a time constant of 153 ms, and showed an off response upon the removal of the stimulus in most cases (Fig. 1d and Supplementary Fig. 7), similar to touch-sensitive mechanoreceptors in *Caenorhabditis elegans* and mice[1,15–17]. Thus, class III dendritic arborization neurons are low-threshold mechanoreceptors for sensing gentle touch.

To identify the mechanotransduction channel that mediates gentle touch, we examined the role of NOMPC, which is important for hearing and mechanotransduction in adult *Drosophila* and larval locomotion[12,18–21]. Our immunostaining of *Drosophila* larvae confirmed high levels of NOMPC expression in the ciliate tips of type I sensory neurons[10,18,22] (Supplementary Fig. 8b), and we found high levels of NOMPC throughout the soma and dendrites of class III dendritic arborization neurons of wild-type but not *nompC* null mutant larvae (Fig. 1f and Supplementary Fig. 8a, c). We further uncovered severe defects of *nompC* null mutant larvae in the behavioural response to gentle touch, which could be rescued by expressing NOMPC in class III neurons (Fig. 1g, h), indicating that NOMPC functions cell autonomously in these neurons to mediate gentle-touch sensation. Moreover, class III neurons of the *nompC* mutant larvae failed to respond to touch stimuli with action potential firing (Fig. 1i) or calcium response (Fig. 1j, k) despite their normal morphology (Supplementary Fig. 9), a defect that could also be rescued by class III neuron expression of NOMPC (Fig. 1i–k). Two *Drosophila* TRP channels, Inactive (IAV) and Nanchung (NAN), are interdependent in their contribution to the hearing of adult *Drosophila*[4]. We found normal mechanosensitivity of class III dendritic arborization neurons in *iav* null mutant larvae (Fig. 1i, k), which showed a normal behavioural response to gentle touch (Fig. 1g, h). Thus, NOMPC, but not IAV, is required for mechanotransduction of larval class III dendritic arborization neurons that sense touch.

We next tested whether touch sensitivity can be conferred to class IV dendritic arborization neurons—which normally do not express NOMPC (Fig. 2a and Supplementary Fig. 10)—by expressing NOMPC ectopically (Fig. 2b). Although class IV neurons with the *ppk*-Gal4 driver alone were insensitive to touch, class IV neurons with NOMPC expression responded to the touch stimulation with a burst of action potentials (Fig. 2c–e) and an increase in internal calcium level (Fig. 2f–h). Thus, not only is NOMPC necessary for mechanosensitivity of class III neurons, but its ectopic expression is sufficient to enable the normally touch-insensitive class IV neurons to respond

[1]Howard Hughes Medical Institute, Departments of Physiology, Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94158, USA.
*These authors contributed equally to this work.

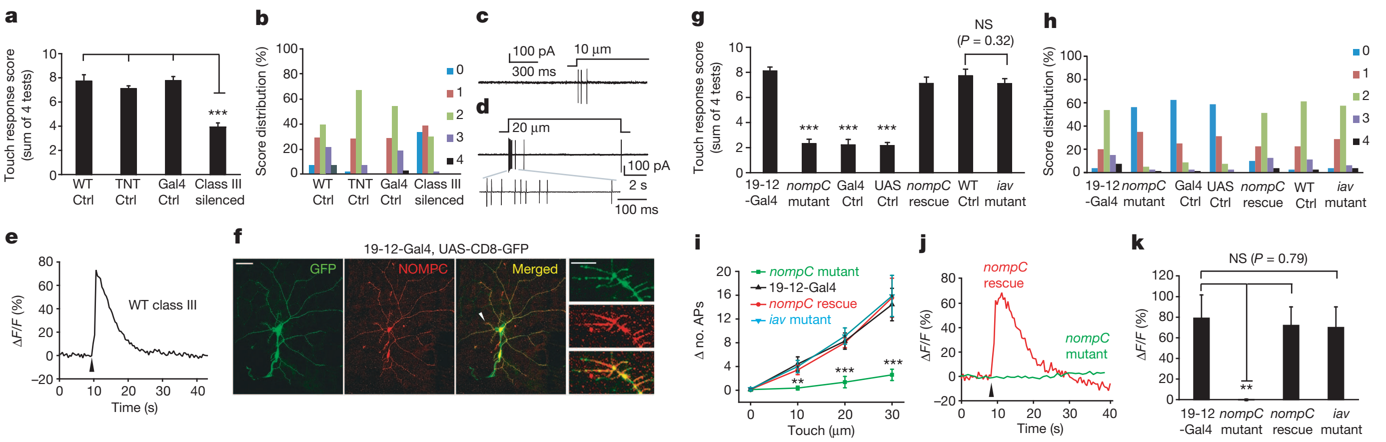


Figure 1 | NOMPC is required for mechanotransduction of class III dendritic arborization neurons that function in gentle-touch sensation of larvae. **a**, Silencing class III neurons with TNT expression impaired larval behavioural responses to gentle touch ($n = 24$–26). WT, wild type. **b**, Score distribution of different genotypes. 0, no response; 1, pause; 2, recoil; 3, single reverse contractile wave; 4, multiple reverse contractions. **c**, Touching the dendrite field of a class III dendritic arborization neuron to cause a 10-µm displacement induced a burst of action potentials. **d**, Class III dendritic arborization neurons responded to prolonged mechanical stimulation, with an on response displaying rapid adaptation as well as an off response. **e**, Internal calcium increase in a class III dendritic arborization neuron in response to a 40-µm touch displacement lasting 300 ms. Arrowhead indicates stimulus onset. **f**, Immunostaining of NOMPC in ddaF and ddaA class III dendritic arborization neurons (labelled by class III-neuron-specific 19-12-Gal4-driven GFP). Arrowhead marks a dendritic segment with NOMPC expression in dendritic spikes, as shown at high magnification on the right. Scale bars: left, 50 µm; right, 5 µm. **g**, *nompC* mutant was defective in gentle-touch sensation; this behavioural phenotype could be rescued by expressing NOMPC in class III dendritic arborization neurons (*nompC* rescue), but not by UAS-NOMPC (UAS Ctrl) or Gal4 (Gal4 Ctrl) alone. *nompC* mutant, Gal4 Ctrl and UAS Ctrl were compared to both 19-12-Gal4 and *nompC* rescue for the significance test. *iav* mutant showed a normal behavioural response to gentle touch compared with wild-type control ($n = 20$–24). **h**, Score distribution of different genotypes. **i**, Summary of action potential firing of class III dendritic arborization neurons in the *nompC* mutant, 19-12-Gal4, *nompC* rescue, and *iav* mutant in response to increasing mechanical stimuli. y axis denotes increase in the number of action potentials (APs) in 1 s after stimulus onset compared to 1 s before stimulus onset. *nompC* mutant showed markedly reduced responses compared to 19-12-Gal4 and *nompC* rescue, whereas the *iav* mutant showed no difference from 19-12-Gal4 ($P > 0.2$; $n = 7$–10). **j**, The calcium response of class III dendritic arborization neurons in *nompC* mutant (green) and *nompC* rescue (red). Arrowhead indicates the onset of the stimulus (40-µm displacement lasting for 300 ms). **k**, Group data of calcium response of class III dendritic arborization neurons with different genotypes ($n = 7$–10). We used unpaired t-test for comparison between two groups, and one-way analysis of variance followed by Tukey's comparison for comparison among three or four groups. NS, not significant. $^{**}P < 0.01$, $^{***}P < 0.001$. All error bars denote ± s.e.m. Genotypes are as follows: for **a**, **b**, wild-type control: w^{1118}. TNT control: *UAS-TNT/+*. Gal4 control: *repo-GAL80/+; 19-12-GAL4/+*. Class III silenced: *repo-GAL80/UAS-TNT; 19-12-GAL4/+*. For **g**, **h**, 19-12-Gal4: *repo-GAL80/+; 19-12-GAL4/+*, *nompC* mutant: $nompC^1/nompC^3$; +/+. Gal4 control: $nompC^1/nompC^3$; *19-12-GAL4/+*. UAS control: $nompC^1/nompC^3$, *UAS-nompC; +/+*. *nompC* rescue: $nompC^1/nompC^3$, *UAS-nompC; 19-12-GAL4/+*. Wild-type control: w^{1118}. *iav* mutant: iav^1. For **i**, 19-12-Gal4: *19-12-GAL4, UAS-CD8-GFP/+*. *nompC* mutant: $nompC^1/nompC^3$; *19-12-GAL4, UAS-CD8-GFP/+*. *nompC* rescue: $nompC^1/nompC^3$, *UAS-nompC; 19-12-GAL4, UAS-CD8-GFP/+*. *iav* mutant: iav^1; *+/+; 19-12-GAL4, UAS-CD8-GFP/+*. For **j**, **k**, 19-12-Gal4: *19-12-GAL4/UAS-GCaMP5*. *nompC* mutant: $nompC^1/nompC^3$; *19-12-GAL4/UAS-GCaMP5*. *nompC* rescue: $nompC^1/nompC^3$, *UAS-nompC; 19-12-GAL4/UAS-GCaMP5*. *iav* mutant: iav^1/y; *+/+; 19-12-GAL4/UAS-GCaMP5*.

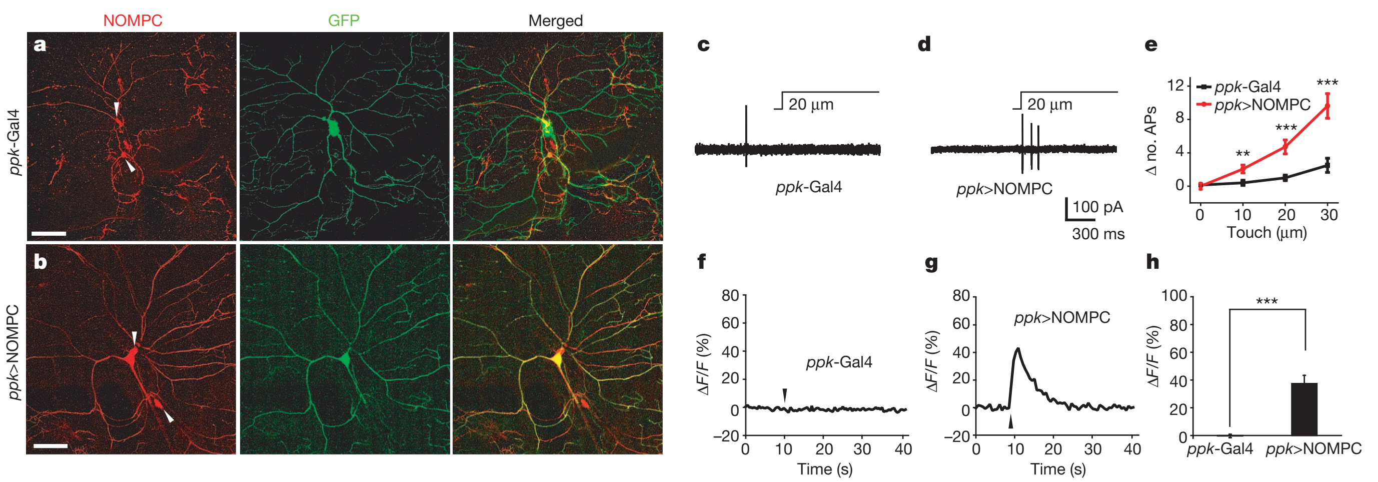


Figure 2 | Ectopic NOMPC expression confers mechanosensitivity to the normally touch-insensitive class IV dendritic arborization neurons. **a**, Class IV neurons showed no detectable NOMPC expression, whereas class III neurons nearby displayed NOMPC staining (indicated by arrowheads). **b**, Class IV neurons with ectopic expression of NOMPC showed strong NOMPC staining. Scale bars in **a** and **b**, 50 µm. **c**, **f**, Class IV neurons did not show detectable electrophysiological (**c**) or calcium (**f**) response to mechanical stimuli. **d**, **g**, Class IV neurons with NOMPC expression displayed touch-induced action potential firing (**d**) and calcium rise upon 40-µm touch displacement lasting 300 ms (**g**). The same mechanical stimuli were used in **f** and **g**. Arrowheads indicate stimulus onset. **e**, **h**, Group data of electrophysiological (**e**) and calcium (**h**) response to mechanical stimulation of class IV neurons from control larvae and larvae with NOMPC expression. *ppk*>NOMPC denotes the UAS-NOMPC/+; *ppk*-Gal4/+ genotype. All error bars denote ± s.e.m. $n = 8$–10. $^{**}P < 0.01$, $^{***}P < 0.001$, unpaired t-test.

to touch. Overexpression of NOMPC also conferred touch sensitivity to the normally touch-insensitive class I dendritic arborization neurons (Supplementary Fig. 11).

Next, we tested whether NOMPC could form mechanosensitive channels in a heterologous expression system. Although there is strong evidence supporting the notion that TRP-4, the *C. elegans* homologue of *Drosophila* NOMPC, is a pore-forming subunit of a native mechanotransduction channel[23], it remains to be tested whether heterologous expression of these channels gives rise to mechanosensitive channels[23]. We found that *Drosophila* S2 cells transfected with NOMPC displayed spontaneous channel openings in whole-cell recordings (Fig. 3a and Supplementary Fig. 12). This current could be blocked reversibly by gadolinium (Gd^{3+}) and SKF-96365 (Fig. 3l and Supplementary Fig. 13a, b), which are known to block many TRP channels. The current could also be blocked by FM1-43, a blocker for mechanotransduction channels on hair cells[24] (Fig. 3n).

We then tested whether NOMPC channels could be gated by mechanical force. First, by using a glass probe driven by a piezo actuator to deliver mechanical stimuli to the cell membrane during whole-cell recording, we observed both on and off responses (Fig. 3b). As expected for mechanically gated channels with fast activation (Fig. 3b), with increasing stimulation strength the current amplitude increased progressively (Fig. 3c) whereas the latency decreased to less than 1.5 ms (Fig. 3d). Gd^{3+} and FM1-43 blocked these mechanosensitive channels, and the channel activity recovered after washout (Fig. 3m, o).

Next, we applied negative pressure through the recording pipette with a high-speed pressure clamp to stretch the membrane of an outside-out excised patch, and found that the NOMPC channel was activated within 2 ms of the pressure application (Fig. 3e). As the pressure was gradually increased, the current amplitude increased progressively (Fig. 3f–h). The pressure-induced NOMPC current was also sensitive to Gd^{3+} and FM1-43 (Supplementary Fig. 13c, d).

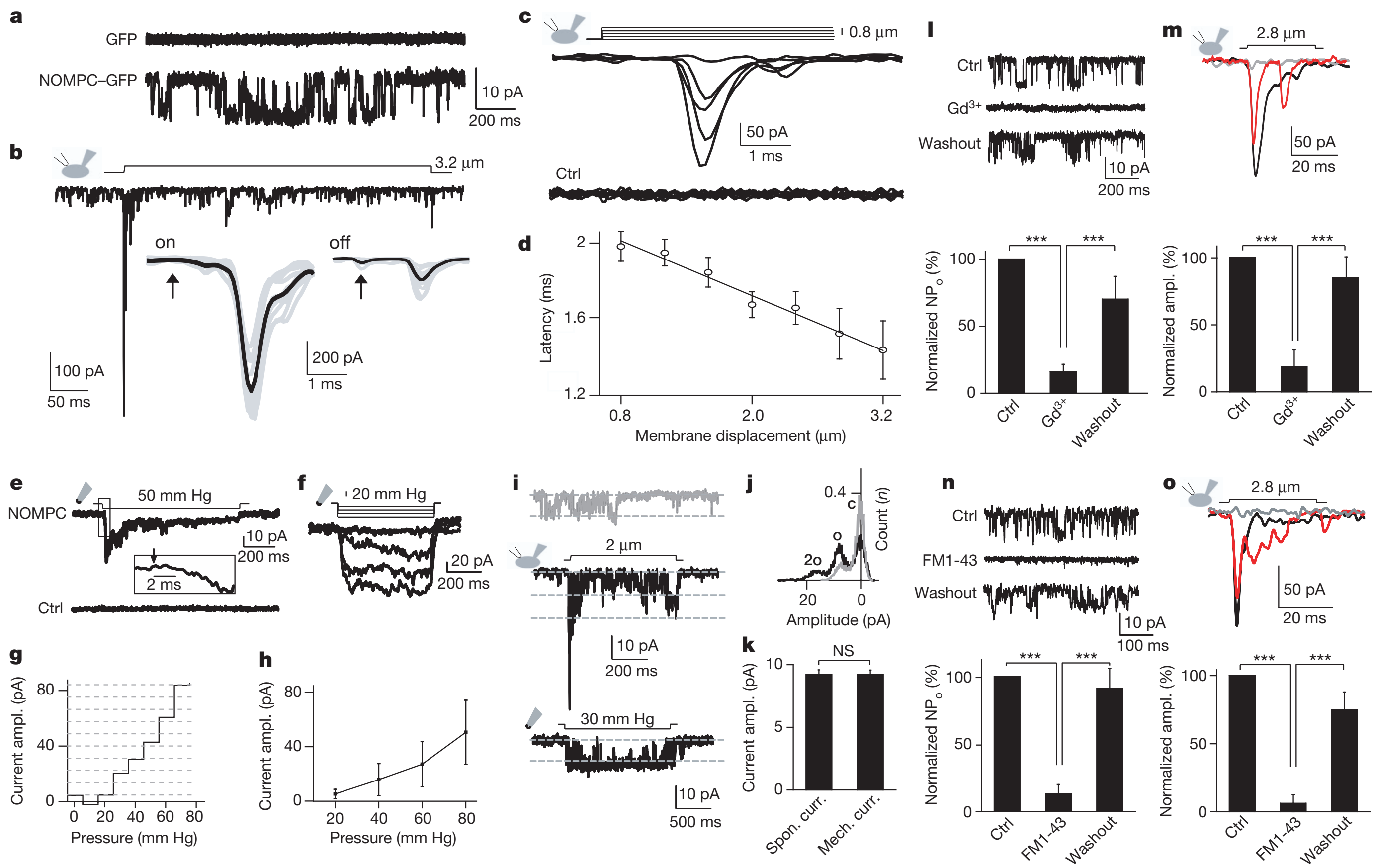


Figure 3 | Heterologous expression of NOMPC in S2 cells yields mechanosensitive channels. **a**, Representative traces of spontaneous channel activities in control and NOMPC–GFP-expressing S2 cells at −60 mV. **b**, Mechanosensitive current triggered by the piezo-actuator-driven probe causing a 3.2-µm displacement of the cell membrane. Insets, the latency is less than 2 ms (arrows indicate the onset and the end of the stimulation for the on and off response, respectively). **c**, The current amplitude (at −60 mV) varied with the size of the displacement (5 trials with 0.4-µm increment, from 0 to 1.6 µm). **d**, The latency of the current response as a function of the mechanical displacement ($n = 7$). Error bars denote ± s.e.m. **e**, Mechanosensitive current of an outside-out patch in response to negative pressure applied with a high-speed pressure clamp. Arrow indicates the pressure onset. **f**, The mechanosensitive current amplitude varied with the pressure intensity (at −60 mV). **g**, Surface plot showing the current amplitude on a single patch increased progressively with increasing negative pressure at −60 mV. **h**, Group data of the dose-dependent curve of pressure-induced current ($n = 7$). Error bars denote ± s.e.m. **i**, Single-channel current (at −60 mV) from spontaneously active channels (top trace) and channels activated by mechanical stimulation via probe displacement (middle trace) and negative pressure on excised patch (bottom trace). **j**, Histogram showing the increase of channel open probability after stimulation. Black denotes displacement-triggered channel activity and grey denotes spontaneous channel activity. c, closed; o, open. **k**, Comparison of the single-channel current amplitude of spontaneously active channels and pressure-induced channels on a patch from the same cell at −60 mV. mech. curr., mechanicosensitive current; spon. curr., spontaneous current. $n = 7$; paired *t*-test; error bars denote ± s.e.m. **l**, Spontaneously active channels in S2 cells expressing NOMPC–GFP were sensitive to 100 µM Gd^{3+} and the block was reversible upon washout ($n = 6$). NP_o indicates open probability. **m**, The displacement-induced mechanosensitive current was blocked by 100 µM Gd^{3+} and then recovered after washout (black, control; grey, Gd^{3+}; red, washout; $n = 4$). **n**, Spontaneously active channels in S2 cells expressing NOMPC–GFP were sensitive to 3 µM FM1-43 and the block was reversible upon washout ($n = 6$). **o**, The displacement-induced mechanosensitive current was blocked by 3 µM FM1-43 and then recovered after washout. Black, control; grey, FM1-43; red, washout; $n = 5$. ***$P < 0.001$, paired *t*-test. All error bars denote ± s.e.m.

By resolving the single-channel current amplitude for mechanosensitive channels (Fig. 3i, j), we found that the single-channel current amplitude for channels activated by negative pressure is identical to that for the spontaneously active channels (Fig. 3i, k). Reversal potential measurements under bi-ionic conditions further revealed that NOMPC channels are permeable not only to monovalent cations such as Na^+, K^+ and Cs^+, but also the divalent cation Ca^{2+} (Supplementary Fig. 14).

A previous study has shown that mutations in several residues of the selective filter of the TRP-4 channel in *C. elegans* change the ion selectivity of the channel[23]. On the basis of the sequence homology between NOMPC and TRP-4 channels, we mutated acidic residues near the putative pore region of NOMPC[23,25,26] (Fig. 4a). Whereas alanine substitution of Asp 1516 had no detectable effects on channel properties including the single-channel conductance (Fig. 4b), replacing Glu 1511 with the basic residue lysine considerably reduced channel activity (Fig. 4c). Alanine substitution of Glu 1511 resulted in a reduction of single-channel conductance measured at −60 mV in Na^+/Cs^+ solutions (Fig. 4b and Supplementary Fig. 15) and altered the shape of current–voltage relationship (*I*–*V*) curves (Fig. 4c and Supplementary Fig. 16). These mutant NOMPC channels showed similar distributions on the cell membrane as wild-type NOMPC (Supplementary Fig. 17). Notably, these mutations had similar effects on the spontaneously active channels and the channels activated by mechanical stimulation. The current response to piezo displacement at −100 mV was reduced in the E1511A mutant, whereas the response was largely eliminated in the E1511K mutant (Fig. 4d, e). The mutations also altered the shape of the *I*–*V* curves for the current induced by both mechanical displacement (Fig. 4e) and negative pressure (Fig. 4f) and shifted the reversal potential of the mechanosensitive channels in Na^+/Cs^+ solutions (Fig. 4g). The wild-type NOMPC channels exhibited a slightly higher permeability (P) to Na^+ than Cs^+ ($P_{Na^+}/P_{Cs^+} = 1.2$), whereas the converse was true for the E1511A mutant ($P_{Na^+}/P_{Cs^+} = 0.7$). These results support the notion that NOMPC is a pore-forming subunit of the mechanotransduction channel.

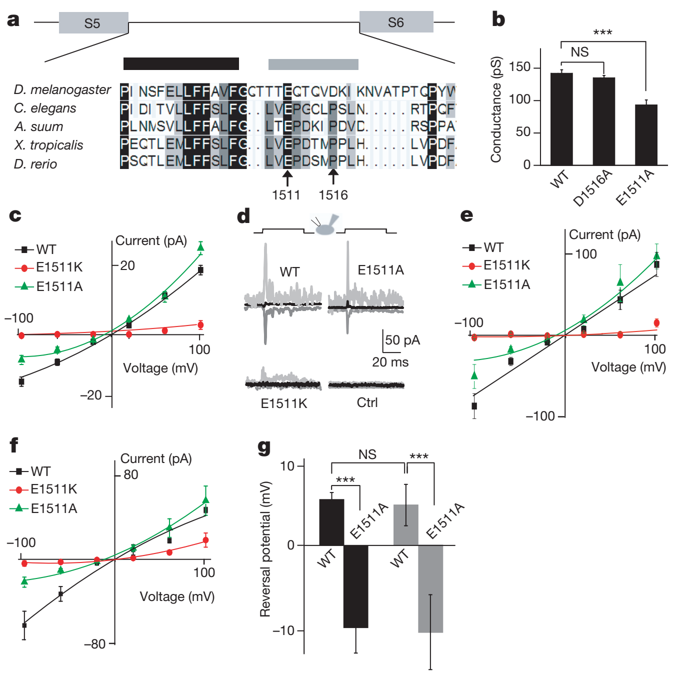


Figure 4 | NOMPC is likely a pore-forming subunit of the mechanotransduction channel. **a**, Sequence alignment of the putative pore region of NOMPC and its homologues, putative pore helix (black bar) and the putative selective filter (grey bar). Arrows mark the negatively charged Glu 1511 and Asp 1516 residues targeted for mutagenesis. *A. suum*, *Ascaris suum*; *X. tropicalis*, *Xenopus tropicalis*; *D. rerio*, *Danio rerio*. **b**, The E1511A mutation altered the single-channel conductance of the spontaneous current determined at −60 mV from whole-cell recording in extracellular Na^+/intracellular Cs^+ solution ($n = 7, 6, 14$ from left to right). $^{***}P < 0.001$, unpaired *t*-test. **c**, The E1511A mutation altered the single-channel *I*–*V* curve of spontaneously active channels in the same Na^+/Cs^+ solutions. The E1511A mutant channel displayed outward rectification and the E1511K mutation greatly reduced channel activity. **d**, Sample traces of wild-type and mutant channels in response to 2.8-μm displacement at −100 (dark grey), 0 (black) and 100 (grey) mV. **e**, The E1511A mutation altered the *I*–*V* curve of displacement (2.8 μm)-triggered currents in the same Na^+/Cs^+ solutions whereas the E1511K mutation largely eliminated the mechanosensitive current at negative voltage. **f**, The E1511A mutation altered the *I*–*V* curve of pressure-elicited currents in the same Na^+/Cs^+ solutions whereas the E1511K mutation largely eliminated the mechanosensitive current (50 mm Hg) at negative voltage. **g**, The E1511A mutation shifted the reversal potential of spontaneously active NOMPC channels (black bars) and NOMPC channels activated by negative pressure (50 mm Hg, grey bars) in the bi-ionic Na^+/Cs^+ conditions ($n = 6, 6, 6, 4$ from left to right). All error bars denote ± s.e.m.

For a channel to be considered a bona fide transducer of mechanical stimuli, it has to meet four criteria[3,27]. NOMPC satisfies all of these criteria: first, it is expressed in class III dendritic arborization neurons that respond to touch; second, it is required for touch-induced excitation of class III neurons; third, ectopic expression of NOMPC endows the normally touch-insensitive dendritic arborization neurons with sensitivity to mechanical stimuli; and fourth, heterologous expression of NOMPC in S2 cells generates non-selective cation channels.

We show that *Drosophila* NOMPC is most likely a pore-forming subunit of a channel that can be activated by mechanical force in a heterologous expression system and is required for mechanotransduction of gentle touch *in vivo*. Together with *Drosophila* PIEZO, which is gated by mechanical force in a heterologous expression system and is required for sensing noxious mechanical stimuli *in vivo*[2,28], these studies reveal that the class III and class IV dendritic arborization neurons that tile the larval body wall use different molecular mechanisms to sense gentle touch and noxious mechanical stimuli, respectively, raising the question of whether different groups of sensory neurons in the skin of other organisms including mammals might also use different molecular mechanisms for sensing these different stimuli types.

METHODS SUMMARY

Behavioural assay. Animals were raised at 25 °C in an incubator. The thoracic segments of the larvae were touched with an eyelash and the behavioural responses were scored as described in ref. 12. 19-12-Gal4 is a marker for class III dendritic arborization neurons and is not expressed in class I, II or IV neurons[29,30].

Electrophysiological recordings. Fillet preparations were made by dissecting third-instar larvae. Action potentials were recorded extracellularly. Whole-cell and outside-out patch recordings of S2 cells were carried out 1–2 days after transfection.

Calcium imaging of dendritic arborization neurons. Calcium imaging of larval dendritic arborization neurons was carried with a newly available genetically coded calcium indicator, GCaMP5.

Mutagenesis of NOMPC channel and S2 cell culture. All point mutations were introduced by site-directed mutagenesis and verified by sequencing of the full-length construct. *Drosophila* S2 cells were transfected using the product protocol.

Mechanical stimulation. A glass probe was driven by a piezo actuator mounted on a micromanipulator to give mechanical stimulation. The movements were triggered and controlled by the piezo amplifier, which was synchronized with the programmed signals from pClamp software. Negative pressure on excised patch was delivered by high-speed pressure clamp.

Immunohistochemistry. In brief, third-instar larvae were dissected in PBS, fixed and exposed to the primary antibody overnight at 4 °C and then the secondary antibody for 2 h at room temperature (25 °C).

Received 14 June; accepted 19 October 2012.
Published online 9 December 2012.

1. Lumpkin, E. A., Marshall, K. L. & Nelson, A. M. The cell biology of touch. *J. Cell Biol.* **191,** 237–248 (2010).
2. Kim, S. E., Coste, B., Chadha, A., Cook, B. & Patapoutian, A. The role of *Drosophila* Piezo in mechanical nociception. *Nature* **483,** 209–212 (2012).

3. Árnadóttir, J. & Chalfie, M. Eukaryotic mechanosensitive channels. *Annu. Rev. Biophys.* **39,** 111–137 (2010).
4. Gong, Z. *et al.* Two interdependent TRPV channel subunits, inactive and Nanchung, mediate hearing in *Drosophila*. *J. Neurosci.* **24,** 9059–9066 (2004).
5. Caldwell, J. C. & Eberl, D. F. Towards a molecular understanding of *Drosophila* hearing. *J. Neurobiol.* **53,** 172–189 (2002).
6. Kamikouchi, A. *et al.* The neural basis of *Drosophila* gravity-sensing and hearing. *Nature* **458,** 165–171 (2009).
7. Hughes, C. L. & Thomas, J. B. A sensory feedback circuit coordinates muscle activity in *Drosophila*. *Mol. Cell. Neurosci.* **35,** 383–396 (2007).
8. Tracey, W. D. Jr, Wilson, R. I., Laurent, G. & Benzer, S. *painless*, a *Drosophila* gene essential for nociception. *Cell* **113,** 261–273 (2003).
9. Yorozu, S. *et al.* Distinct sensory representations of wind and near-field sound in the *Drosophila* brain. *Nature* **458,** 201–205 (2009).
10. Liang, X., Madrid, J., Saleh, H. S. & Howard, J. NOMPC, a member of the TRP channel family, localizes to the tubular body and distal cilium of *Drosophila* campaniform and chordotonal receptor cells. *Cytoskeleton* **68,** 1–7 (2011).
11. Song, W., Onishi, M., Jan, L. Y. & Jan, Y. N. Peripheral multidendritic sensory neurons are necessary for rhythmic locomotion behavior in *Drosophila* larvae. *Proc. Natl Acad. Sci. USA* **104,** 5199–5204 (2007).
12. Kernan, M., Cowan, D. & Zuker, C. Genetic dissection of mechanosensory transduction: mechanoreception-defective mutations of *Drosophila*. *Neuron* **12,** 1195–1206 (1994).
13. Grueber, W. B., Jan, L. Y. & Jan, Y. N. Tiling of the *Drosophila* epidermis by multidendritic sensory neurons. *Development* **129,** 2867–2878 (2002).
14. Zhong, L., Hwang, R. Y. & Tracey, W. D. Pickpocket is a DEG/ENaC protein required for mechanical nociception in *Drosophila* larvae. *Curr. Biol.* **20,** 429–434 (2010).
15. Geffeney, S. L. & Goodman, M. B. How we feel: ion channel partnerships that detect mechanical inputs and give rise to touch and pain perception. *Neuron* **74,** 609–619 (2012).
16. Li, L. *et al.* The functional organization of cutaneous low-threshold mechanosensory neurons. *Cell* **147,** 1615–1627 (2011).
17. Delmas, P., Hao, J. & Rodat-Despoix, L. Molecular mechanisms of mechanotransduction in mammalian sensory neurons. *Nature Rev. Neurosci.* **12,** 139–153 (2011).
18. Cheng, L. E., Song, W., Looger, L. L., Jan, L. Y. & Jan, Y. N. The role of the TRP channel NompC in *Drosophila* larval and adult locomotion. *Neuron* **67,** 373–380 (2010).
19. Effertz, T., Wiek, R. & Gopfert, M. C. NompC TRP channel is essential for *Drosophila* sound receptor function. *Curr. Biol.* **21,** 592–597 (2011).
20. Eberl, D. F., Hardy, R. W. & Kernan, M. J. Genetically similar transduction mechanisms for touch and hearing in *Drosophila*. *J. Neurosci.* **20,** 5981–5988 (2000).
21. Walker, R. G., Willingham, A. T. & Zuker, C. S. A *Drosophila* mechanosensory transduction channel. *Science* **287,** 2229–2234 (2000).
22. Lee, J., Moon, S., Cha, Y. & Chung, Y. D. *Drosophila* TRPN(=NOMPC) channel localizes to the distal end of mechanosensory cilia. *PLoS ONE* **5,** e11012 (2010).
23. Kang, L., Gao, J., Schafer, W. R., Xie, Z. & Xu, X. Z. C. *elegans* TRP family protein TRP-4 is a pore-forming subunit of a native mechanotransduction channel. *Neuron* **67,** 381–391 (2010).
24. Meyers, J. R. *et al.* Lighting up the senses: FM1-43 loading of sensory cells through nonselective ion channels. *J. Neurosci.* **23,** 4054–4065 (2003).
25. Moran, M. M., Xu, H. & Clapham, D. E. TRP ion channels in the nervous system. *Curr. Opin. Neurobiol.* **14,** 362–369 (2004).
26. Venkatachalam, K. & Montell, C. TRP channels. *Annu. Rev. Biochem.* **76,** 387–417 (2007).
27. Christensen, A. P. & Corey, D. P. TRP channels in mechanosensation: direct or indirect activation? *Nature Rev. Neurosci.* **8,** 510–521 (2007).
28. Coste, B. *et al.* Piezo proteins are pore-forming subunits of mechanically activated channels. *Nature* **483,** 176–181 (2012).
29. Xiang, Y. *et al.* Light-avoidance-mediating photoreceptors tile the *Drosophila* larval body wall. *Nature* **468,** 921–926 (2010).
30. Rumpf, S., Lee, S. B., Jan, L. Y. & Jan, Y. N. Neuronal remodeling and apoptosis require VCP-dependent degradation of the apoptosis inhibitor DIAP1. *Development* **138,** 1153–1160 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Zuker, U. Heberlein, G. Rubin, T. Lee and R. Bodmer for fly lines, J. Howard for the NOMPC antibody and L. Looger for GCaMP5. We thank D. Minor and S. Bagriantsev for assistance with the high-speed pressure clamp. We thank S. Younger, S. Barbel and T. Cheng for technical support, W. P. Ge, M. P. Klassen, P. Fan, E. K. Unger and C. J. Peters for reading the manuscript, and members of the Jan laboratory for discussion. Z.Y. and Y.X. are recipients of the Long-Term Fellowship from the Human Frontier Science Program. This work was supported by National Institutes of Health grants (R37NS040929 and 5R01MH084234) to Y.N.J. L.Y.J. and Y.N.J. are investigators of the Howard Hughes Medical Institute.

Author Contributions Z.Y. and W.Z. initiated the project, designed and conducted most experiments; Y.N.J. and L.Y.J. supervised the project and provided guidance throughout. L.E.C. and W.Z. made the NOMPC mutants. Y.H., Y.X. and S.M. assisted with part of the behavioural experiments and immunostaining. D.G. helped set up the piezo actuator system. Z.Y., W.Z., L.Y.J. and Y.N.J. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.N.J. (yuhnung.jan@ucsf.edu).

doi:10.1038/nature11689

Metabolic control of adult neural stem cell activity by Fasn-dependent lipogenesis

Marlen Knobloch[1,2]*, Simon M. G. Braun[1,2]*, Luis Zurkirchen[2], Carolin von Schoultz[2], Nicola Zamboni[3], Marcos J. Araúzo-Bravo[4], Werner J. Kovacs[2], Özlem Karalay[2], Ueli Suter[2], Raquel A. C. Machado[1,2], Marta Roccio[5], Matthias P. Lutolf[5], Clay F. Semenkovich[6] & Sebastian Jessberger[1,2]

Mechanisms controlling the proliferative activity of neural stem and progenitor cells (NSPCs) have a pivotal role to ensure life-long neurogenesis in the mammalian brain[1]. How metabolic programs are coupled with NSPC activity remains unknown. Here we show that fatty acid synthase (Fasn), the key enzyme of *de novo* lipogenesis[2], is highly active in adult NSPCs and that conditional deletion of *Fasn* in mouse NSPCs impairs adult neurogenesis. The rate of *de novo* lipid synthesis and subsequent proliferation of NSPCs is regulated by *Spot14*, a gene previously implicated in lipid metabolism[3–5], that we found to be selectively expressed in low proliferating adult NSPCs. Spot14 reduces the availability of malonyl-CoA[6], which is an essential substrate for Fasn to fuel lipogenesis. Thus, we identify here a functional coupling between the regulation of lipid metabolism and adult NSPC proliferation.

Adult neurogenesis, the generation of new neurons throughout adulthood in discrete brain regions, is critically involved in certain forms of cognition and has been associated with neuropsychiatric diseases such as depression and epilepsy[1,7]. Recently, mechanisms regulating cell metabolism have been shown to control haematopoietic stem cell activity as well as proliferation of certain cancer cells[8–11]. Whether regulation of cell metabolism governs cell proliferation in mammalian NSPCs remains unknown. Therefore, we here studied the role of *de novo* lipogenesis and its key enzyme Fasn[11,12] in the context of adult NSPC biology.

Fasn messenger RNA and protein were expressed in the two main neurogenic areas of the adult brain, the subgranular zone (SGZ) of the dentate gyrus and the subventricular zone (SVZ) lining the lateral ventricles (Fig. 1a and Supplementary Fig. 1a–h). Fasn expression was high in proliferating NSPCs but downregulated *in vitro* after differentiation (Supplementary Fig. 2a). Proliferating NSPCs showed high levels of Fasn enzymatic activity relative to other dividing neural cells, such as non-myelinating Schwann cells, or to NSPC-derived differentiated progeny (Supplementary Fig. 2b, c). High levels of Fasn activity were not due to impaired uptake of fatty acids, as NSPCs were competent to pick up free fatty acids in substantial amounts (Supplementary Fig. 2d). Mass-spectrometry-based metabolomics of dividing NSPCs, proliferating non-myelinating Schwann cells and differentiated progeny of NSPCs revealed striking differences in the metabolic profile of these discrete cell populations, confirming that NSPCs are in a distinct metabolic state (Fig. 1b).

Pharmacological inhibition of Fasn activity using two independent inhibitors of Fasn (orlistat and cerulenin) reduced proliferation of rodent NSPCs in a dose-dependent manner (Fig. 1c and Supplementary Fig. 3a, c, e) without inducing substantial amounts of cell death after short-term treatment (Supplementary Fig. 3b, d). In contrast, proliferation of non-myelinating Schwann cells was not affected by Fasn inhibition (Supplementary Fig. 3f). However, prolonged inhibition of Fasn enhanced apoptotic cell death, suggesting that Fasn activity is also required for NSPC maintenance (Supplementary Fig. 3d). To genetically delete *Fasn*, we isolated NSPCs from mice carrying floxed alleles of the *Fasn* gene (*Fasn*$^{flox/flox}$)[12]. Retrovirus-based Cre-recombinase-mediated excision of *Fasn* phenocopied the effects of pharmacological Fasn inhibition and reduced proliferation of NSPCs (Fig. 1d and Supplementary Fig. 4a–c).

We next conditionally deleted *Fasn* specifically in adult NSPCs *in vivo* by crossing *Fasn*$^{flox/flox}$ mice with mice harbouring tamoxifen-inducible nestin-driven Cre recombinase and yellow fluorescent protein (YFP) reporter alleles in the ROSA locus (Fasn-cKO (conditional knockout))[12,13]. When analysed 40 days after the last tamoxifen injection, the number of recombined, YFP-expressing cells was dramatically reduced in both neurogenic areas in Fasn-cKO mice compared to their respective controls (Fig. 1e, f and Supplementary Fig. 4d). Phenotyping of YFP-labelled cells in the dentate gyrus revealed that the defects observed after conditional *Fasn* deletion manifested themselves early in the neurogenic lineage with a reduction in YFP-labelled cells occurring already at the stage of radial NSPCs and resulting in a dramatic loss of newly generated neurons (Fig. 1g, h and Supplementary Fig. 5a, b). These results show that adult NSPCs require high levels of *de novo* lipogenesis for proper neurogenesis.

We next sought to identify genes that could regulate *de novo* lipogenesis in adult NSPCs and consequently alter proliferation of NSPCs. One gene that has previously been implicated as a context-dependent modulator of *de novo* lipogenesis is *Spot14* (also known as thyroid hormone responsive protein (*Thrsp*))[3–5]. *In situ* hybridization revealed that *Spot14* mRNA expression was highly confined to the two neurogenic areas of the adult brain (Fig. 2a and Supplementary Fig. 6a–c, e). To study Spot14-expressing cells in more detail we used a bacterial artificial chromosome (BAC) transgenic mouse driving green fluorescent protein (GFP) under the control of the regulatory elements of the *Spot14* gene (Spot14-GFP). In agreement with the mRNA expression, Spot14-GFP-expressing cells (hereafter called Spot14$^+$ cells) were largely restricted to the SGZ with only scattered expression outside the dentate gyrus, unlike previously described NSPC markers such as nestin, which also labels a substantial number of cells outside the adult neurogenic niches (Fig. 2a and Supplementary Fig. 6d, f). Spot14$^+$ cells in the dentate gyrus showed a radial and non-radial cell morphology[14], and co-labelled with NSPC markers such as Sox2 ($92.7 \pm 15\%$), glial fibrillary acidic protein (Gfap; $85.3 \pm 4.2\%$ in radial cells) and nestin ($73.4 \pm 5.7\%$ in radial cells) (Fig. 2b and Supplementary Fig. 6g, h)[14]. Of all Spot14$^+$ cells, only $9.36 \pm 1.1\%$ were dividing. Non-radial Spot14$^+$ cells showed higher proliferation rates ($7.86 \pm 1.0\%$) compared to radial Spot14$^+$ cells ($1.7 \pm 0.5\%$), as measured by Ki67 co-labelling (Fig. 2b and Supplementary Fig. 6i), indicating that the majority of Spot14$^+$ cells are not dividing under resting conditions.

[1]Brain Research Institute, Faculty of Medicine, University of Zurich, 8057 Zurich, Switzerland. [2]Institute of Molecular Health Sciences, Department of Biology, Swiss Federal Institute of Technology (ETH) Zurich, 8093 Zurich, Switzerland. [3]Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, 8093 Zurich, Switzerland. [4]Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, 48149 Muenster, Germany. [5]Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. [6]Washington University School of Medicine, Division of Endocrinology, Metabolism & Lipid Research, St. Louis, Missouri 63110, USA.
*These authors contributed equally to this work.

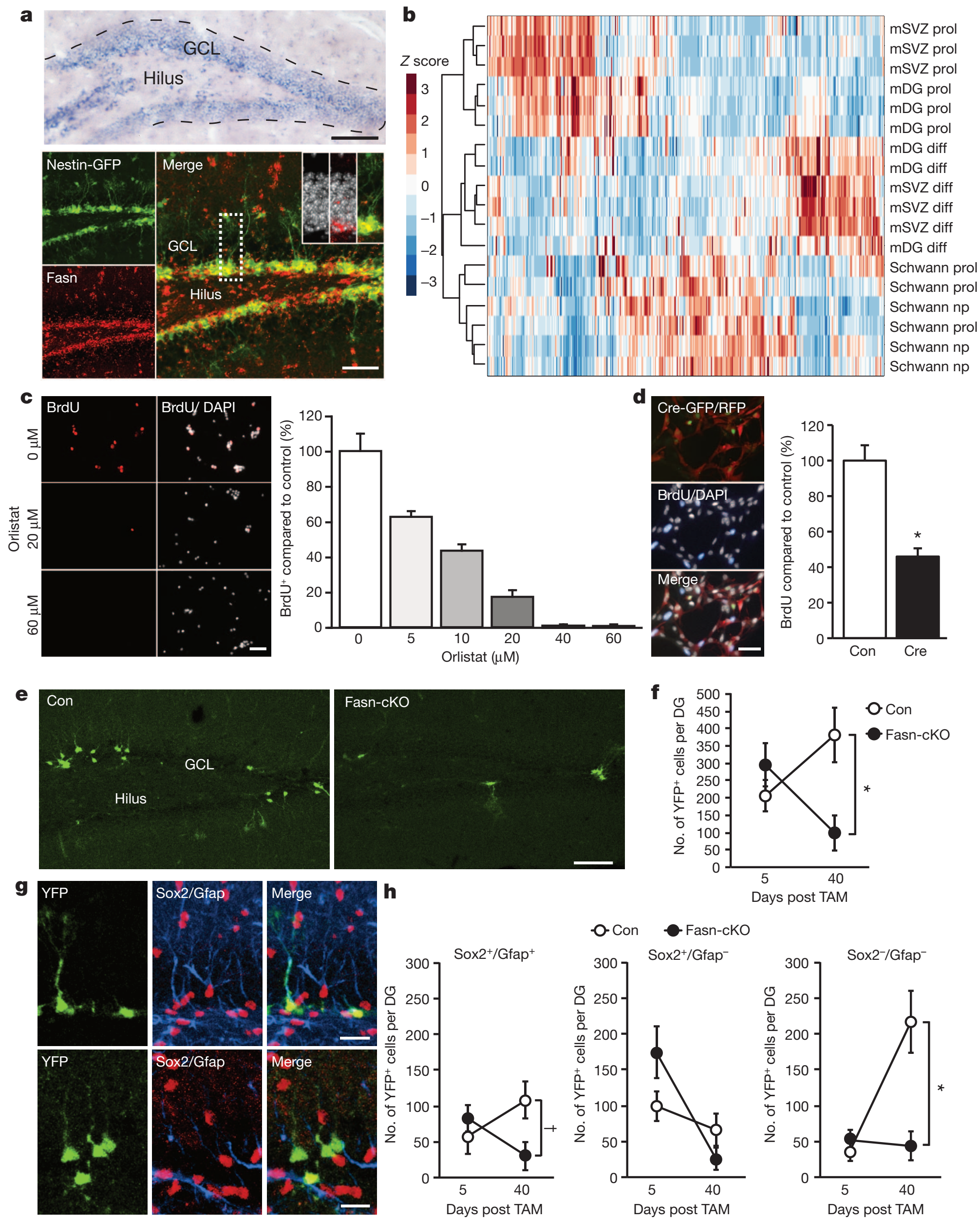


Figure 1 | NSPCs are in a distinct metabolic state and require Fasn-dependent lipogenesis for proliferation. **a**, *In situ* hybridization shows expression of *Fasn* within the neurogenic dentate niche. Fasn protein is expressed in nestin-GFP^{+} cells. Insets show high-power views of Fasn-expressing cells with 4′,6-diamidino-2-phenylindole (DAPI; grey). GCL, granule-cell layer. **b**, Metabolomics show a distinct metabolic state of proliferating dentate gyrus and SVZ NSPCs (mDG prol and mSVZ prol) compared to proliferating Schwann cells (Schwann prol), non-proliferating Schwann cells (Schwann nprol), and differentiated progeny of hippocampal and SVZ NSPCs (mDG diff and mSVZ diff). Note the close clustering of distinct cell populations. **c**, Pharmacological inhibition of Fasn using orlistat leads to a dose-dependent reduction of hippocampal NSPC proliferation. BrdU, 5-bromodeoxyuridine. **d**, Genetic deletion of *Fasn* decreases proliferation of adult hippocampal NSPCs. Con, control; Cre, transduction of Cre-expressing virus to delete Fasn. Left panel, cells that were transduced with a red fluorescent protein (RFP)-expressing control virus and Cre-GFP-expressing retrovirus, and pulsed with BrdU. **e**, Conditional *Fasn* deletion in NSPCs reduces hippocampal neurogenesis. Shown are YFP-expressing cells in controls and mutant mice (Fasn-cKO). **f**, Quantification of YFP-labelled cells in the dentate gyrus (DG) 5 days and 40 days after tamoxifen (TAM)-induced recombination. **g**, Phenotyping of recombined, YFP-labelled cells reveals that conditional Fasn deletion reduces the number of radial glia-like NSPCs expressing Sox2 and Gfap 40 days after tamoxifen and leads to a substantial decrease in differentiated YFP-labelled cells. **h**, Quantification of phenotypes after conditional *Fasn* deletion. Error bars represent mean ± s.e.m. Scale bars represent 100 µm (**a** top panel, **e**), 40 µm (**a** bottom panel), 50 µm (**c**, **d**) and 20 µm (**g**). $^{*}P < 0.05$, $^{\dagger}P = 0.055$.

Confirming the relative quiescence of Spot14^{+} cells, we found that out of all Ki67^{+} cells within the dentate gyrus, only 5.1 ± 0.7% were positive for Spot14-GFP, whereas in the nestin-GFP mouse, 46.6 ± 3.7% of all Ki67^{+} cells were GFP positive (Supplementary Fig. 6j).

To show directly that cells expressing Spot14 possess neurogenic stem-cell potential in the adult brain we generated a BAC transgenic mouse expressing tamoxifen-inducible Cre under the regulatory elements of the *Spot14* gene (Spot14-CreER^{T2}). Spot14-CreER^{T2} mice were crossed with a ROSA reporter line expressing YFP upon Cre-based recombination, enabling *in vivo* lineage tracing of Spot14-expressing cells (Supplementary Fig. 7a). Soon after tamoxifen injections (after 10 days) we found YFP labelling of radial (Sox2^{+}Gfap^{+}) and non-radial (Sox2^{+}Gfap^{-}) NSPCs in the dentate gyrus as well as very few astrocytes (Fig. 2c, d). At later time points after tamoxifen

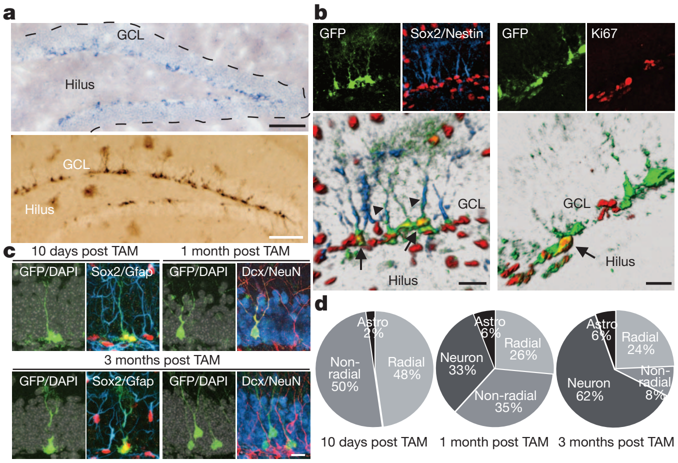


Figure 2 | Spot14 is expressed in adult hippocampal NSPCs. **a**, *In situ* hybridization shows restricted expression of *Spot14* in the SGZ of the dentate gyrus (top panel). Enhanced GFP expression under the regulatory elements of the *Spot14* gene labels cells in the SGZ (bottom panel). **b**, Spot14-GFP co-labels with NSPC markers Sox2 and nestin. Arrows, GFP and Sox2 co-labelled cells; arrowheads, GFP and nestin co-labelled radial glia-like processes (bottom-left panel). A fraction of Spot14-GFP cells divide. Arrow, a Ki67-positive, Spot14-GFP co-labelled cell (bottom-right panel). **c**, YFP expression in radial and non-radial NSPCs labelled by Sox2 and Gfap 10 days after tamoxifen (TAM) injections in Spot14-CreER^{T2} mice. One month after tamoxifen injections, YFP-expressing cells co-label with Dcx, and 3 months after tamoxifen injections these cells co-label with the mature neuronal marker NeuN (bottom-right panel). Note also that radial NSPCs were labelled with YFP 3 months after tamoxifen injections (bottom-left panel). DAPI is shown in grey. **d**, Quantification of cellular phenotypes generated 10 days, 1 month and 3 months after tamoxifen injections in Spot14-CreER^{T2} mice. Astro, astrocytes. Scale bars represent 100 μm (**a**) and 20 μm (**b**, **c**).

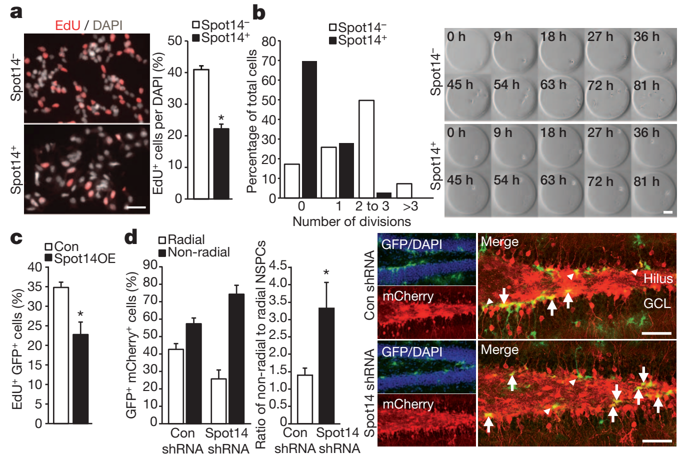


Figure 3 | Spot14 regulates proliferative activity of adult NSPCs. **a**, Hippocampal Spot14^{+} cells are less proliferative than Spot14^{-} NSPCs. **b**, Time-lapse imaging of single hippocampal NSPCs cultured in hydrogel-based microwells reveals reduced proliferation of Spot14^{+} cells compared to Spot14^{-} NSPCs on a single-cell level. Shown are snapshots every 9 h for Spot14^{-} cells (top panels) and Spot14^{+} (bottom panels), and the quantification of the number of cell divisions over 81 h. **c**, Retroviral overexpression of Spot14 (Spot14OE) reduces proliferation of hippocampal NSPCs. **d**, shRNA-mediated knockdown of *Spot14* (Spot14 shRNA) leads to a shift from radial towards non-radial Spot14^{+} NSPCs compared to a non-targeting, shRNA-expressing lentivirus (Con shRNA). Arrows, non-radial NSPCs; arrowheads, radial NSPCs. Error bars represent mean ± s.e.m. Scale bars represent 40 μm (**a**) and 100 μm (**b**, **d**). *$P < 0.05$.

injections (1 month and 3 months later) we found YFP labelling of new granule cells expressing doublecortin (Dcx) and the pan-neuronal marker NeuN (neuronal nuclei; also known as RNA binding protein Fox1 homologue 3) besides labelling of radial and non-radial NSPCs, both in the dentate gyrus and SVZ (Fig. 2c, d and Supplementary Fig. 7b–d). These findings indicate that Spot14-expressing cells constitute neurogenic NSPCs within the adult brain.

To further characterize Spot14^{+} cells on functional and molecular levels, we prospectively isolated NSPCs from Spot14-GFP transgenic mice and used FACS to separate Spot14^{+} and Spot14^{-} NSPCs (Supplementary Fig. 8a, b). Both Spot14^{+} and Spot14^{-} NSPCs expressed Sox2 and nestin (data not shown), self-renewed over an extended period of time *in vitro* and gave rise to neuronal and glial cells, indicating multipotency (Supplementary Fig. 8c–e). However, we observed reduced proliferation of Spot14^{+} cells compared to Spot14^{-} cells using monolayer cultures and clonal time-lapse imaging (Fig. 3a, b and Supplementary Videos 1–4)[15], suggesting that the low proliferative activity of Spot14^{+} cells within the dentate niche is maintained during culture conditions, a finding that was also supported by gene expression profiling (Supplementary Fig. 9a, b and Supplementary Tables 1–6).

Given the low proliferation rate of Spot14^{+} cells *in vivo* and *in vitro* we tested whether manipulation of *Spot14* expression levels impacts NSPC proliferation. Retrovirus-mediated overexpression of *Spot14* reduced proliferation of cultured NSPCs (Fig. 3c and Supplementary Fig. 10a). To test for a function of *Spot14* in NSPC activity *in vivo*, we injected lentiviruses expressing short hairpin RNAs (shRNAs) and the fluorescent label mCherry into Spot14-GFP reporter mice. Knockdown of *Spot14* mediated by shRNAs led to a shift in Spot14^{+} cells towards non-radial, more-proliferative NSPCs, compared to control cells expressing non-targeting shRNAs (Fig. 3d and Supplementary Fig. 10b).

We next sought to identify a mechanism by which *Spot14* controls NSPC proliferation. Spot14 has been shown to interact with Mig12 (Mid1-interacting protein 1), an activator of acetyl-coenzyme A (CoA)-carboxylase (Acc), the enzyme that catalyses the production of malonyl-CoA, which is a substrate for Fasn to generate new palmitate[16]; however, heterodimerization with Spot14 impairs the ability of Mig12 to activate Acc, leading to decreased Acc activity[6]. We found expression of Fasn in Spot14^{+} cells and high levels of Mig12 in adult NSPCs (Fig. 4a, b and Supplementary Fig. 11a–d). In addition, Spot14 and Mig12 were both localized to the cytoplasm and co-immunoprecipitated (Supplementary Fig. 11e, f), suggesting that an interaction between Spot14 and Mig12 takes place in adult NSPCs. Retroviral overexpression of Mig12 rescued the Spot14-induced decrease in NSPC proliferation, indicating that a Spot14–Mig12 interaction is involved in the regulation of NSPC proliferation (Supplementary Fig. 11g). Furthermore, virus-mediated knockdown of Mig12 reduced proliferation of NSPCs within the adult hippocampus *in vivo* and *in vitro*, as measured by EdU (5-ethynyl-2′-deoxyuridine) pulse labelling (Fig. 4c and Supplementary Fig. 11h, i). Using targeted liquid chromatography tandem mass spectrometry we found that Spot14 overexpression led to a significant drop in malonyl-CoA levels without affecting expression levels of Fasn protein (Fig. 4d and Supplementary Fig. 10c), suggesting that Spot14 decreases malonyl-CoA levels and as a consequence lowers substrate availability for Fasn to fuel *de novo* lipogenesis. To directly assess levels of *de novo* lipogenesis upon Spot14 overexpression we measured total *de novo* lipid production by quantifying incorporation of radioactively labelled ^{14}C-acetate and ^{14}C-labelled glucose into lipids. We found that Spot14 overexpression significantly decreased levels of lipid synthesis (Fig. 4e). These data support previous reports that Spot14 can act as a context-dependent modulator of *de novo* lipogenesis, either as an inhibitor[3] (as shown here for adult NSPCs) or activator of lipid synthesis[4,5,17]. In support of an essential role for Fasn and Spot14, we found that genetic deletion of *Fasn* or overexpression of Spot14 (both leading to reduced *de novo* lipogenesis) induced a shift in the metabolic profile of NSPCs, as determined by principal component analyses of

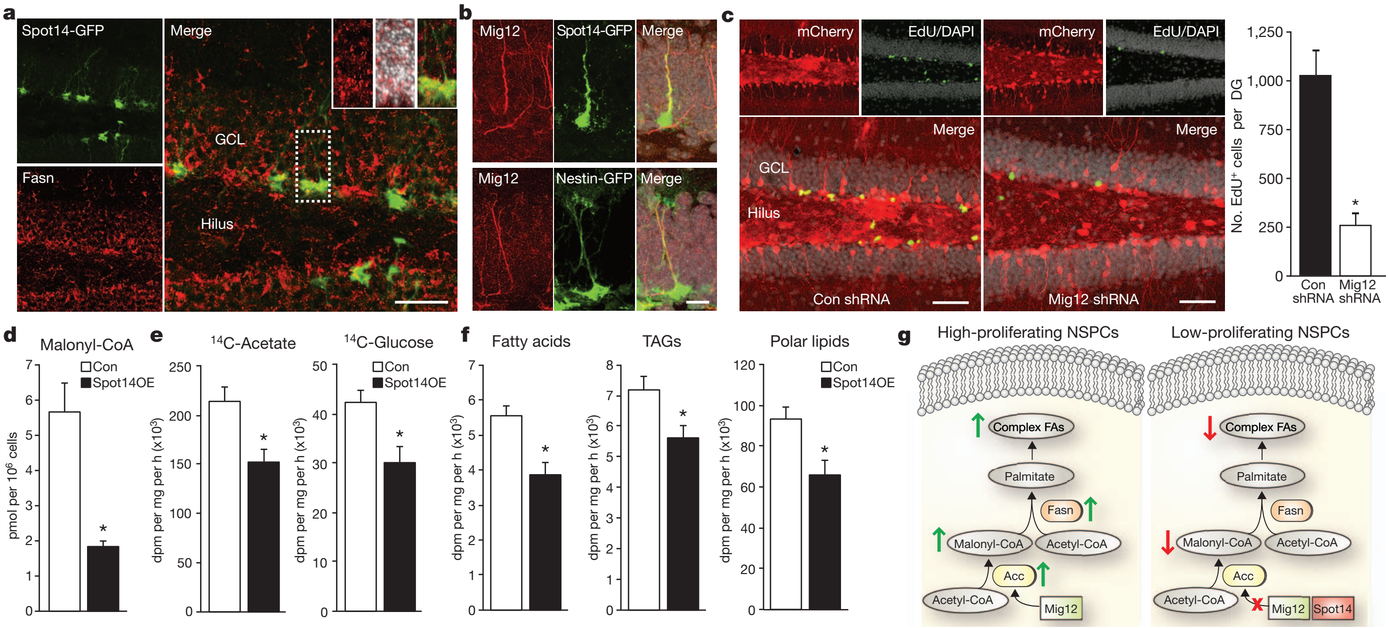


Figure 4 | Spot14 modulates malonyl-CoA levels and *de novo* lipogenesis. **a**, Fasn protein is expressed in Spot14$^+$ cells. Insets show high-power views of a Spot14-GFP and Fasn co-labelled cell. **b**, Mig12 is expressed in adult NSPCs. **c**, shRNA-mediated knockdown of Mig12 *in vivo* (Mig12 shRNA, right panel) reduces proliferation of hippocampal NSPCs. **d**, Retroviral overexpression of Spot14 reduces malonyl-CoA levels in NSPCs. dpm, disintegrations per minute; pmol, picomol. **e**, Radioactive tracing using ^{14}C-acetate and ^{14}C-glucose shows a decrease in *de novo* lipogenesis after Spot14 overexpression. **f**, Lipid separation by thin-layer chromatography after ^{14}C-acetate labelling shows a reduction in *de novo* synthesized free fatty acids, TAGs (triacylglycerols) and polar lipids in Spot14-overexpressing hippocampal NSPCs. **g**, Scheme showing the proposed mechanism: high-proliferating NSPCs require Fasn-dependent lipogenesis. Spot14 downregulates NSPC proliferation by reducing the Fasn-dependent generation of new complex fatty acids (FAs). Error bars represent mean ± s.e.m. Scale bars represent 40 µm (**a**), 20 µm (**b**) and 50 µm (**c**). *$P < 0.05$.

metabolomics experiments (Supplementary Fig. 12). We next analysed the fate of newly synthesized lipids using thin-layer chromatography (TLC) and found that the majority of them (91.9%; Fig. 4g) ended up in the membrane lipid-containing fractions, suggesting that NSPCs require a large part of new lipids for structural membranes.

We identify here a novel mechanism governing adult neurogenesis by showing that adult NSPCs require Fasn-dependent lipogenesis for proliferation and that *Spot14*, a gene highly enriched in low proliferating NSPCs, downregulates NSPC proliferation by reducing lipid synthesis (Fig. 4g). Thus, Spot14 seems to act as a molecular brake on Fasn-dependent lipogenesis, potentially also integrating other signals, such as thyroid hormone-regulated pathways, to ensure relative quiescence of adult NSPCs. However, future studies will have to further dissect and characterize the exact cellular mechanisms and NSPC-subtype during which high levels of lipogenesis are required for proper proliferation.

The targeting of metabolic pathways may hold potential for novel therapeutic approaches to treat diseases associated with failing adult neurogenesis, such as major depression or age-related cognitive decline.

METHODS SUMMARY

Mouse strains used were Spot14-GFP (Mutant Mouse Regional Resource Centres; MMRRC), nestin-GFP[18], Fasn floxed[12], nestin-CreER^{T2}-R26YFP[13] and Spot14-CreER^{T2}. The Spot14-CreER^{T2} BAC transgenic line was generated according to a published protocol[19]. All animal experiments were approved by the veterinary office of the Canton of Zurich, Switzerland. Cre-mediated recombination was induced by five consecutive intraperitoneal injections of tamoxifen (180 mg kg^{-1}) at the age of 6 to 8 weeks. Retro- and lentiviral constructs, their production and stereotactic injections were carried out as described previously[20]. Brain tissue was sectioned and stained using methods published previously[20]. Murine NSPCs were cultured in DMEM with Ham's F12, supplemented with N2 supplement plus EGF, FGF and heparin. Fasn enzyme activity was determined using an assay described previously[12]. Spot14$^+$ and Spot14$^-$ cells were live sorted and collected for RNA extraction, and gene expression analysis was then carried out using Affymetrix GeneChip arrays. To measure *de novo* lipid production in NSPCs overexpressing Spot14 and controls, cells were incubated with [1-^{14}C]acetate or D-[U-^{14}C]glucose and the organic phases of the lipid extracts were transferred to scintillation vials, and ^{14}C and ^{3}H counts were measured. Metabolomics samples were analysed by ion pairing–reverse phase liquid chromatography tandem mass spectrometry on a Waters Acquity UHPLC coupled to a Thermo TSQ Quantum Ultra triple quadrupole instrument. Statistical analyses were carried out using two-tailed unpaired *t*-tests, non-parametric Mann–Whitney *U*-tests, the Kolmogorov–Smirnov test, or analysis of variance (ANOVA) followed by Fisher's post hoc test if appropriate. Significance levels were set at $P < 0.05$ ($P < 0.01$ for array analyses). A detailed methods section is available in the Supplementary Information.

Received 31 October 2011; accepted 19 October 2012.

Published online 2 December 2012; corrected online 9 January 2013 (see full-text HTML version for details).

1. Zhao, C., Deng, W. & Gage, F. H. Mechanisms and functional implications of adult neurogenesis. *Cell* **132,** 645–660 (2008).
2. Maier, T., Jenni, S. & Ban, N. Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **311,** 1258–1262 (2006).
3. Zhu, Q. *et al.* Spot 14 gene deletion increases hepatic *de novo* lipogenesis. *Endocrinology* **142,** 4363–4370 (2001).
4. Moncur, J. T., Park, J. P., Memoli, V. A., Mohandas, T. K. & Kinlaw, W. B. The "Spot 14" gene resides on the telomeric end of the 11q13 amplicon and is expressed in lipogenic breast cancers: implications for control of tumor metabolism. *Proc. Natl Acad. Sci. USA* **95,** 6989–6994 (1998).
5. Campbell, M. C., Anderson, G. W. & Mariash, C. N. Human spot 14 glucose and thyroid hormone response: characterization and thyroid hormone response element identification. *Endocrinology* **144,** 5242–5248 (2003).
6. Colbert, C. L. *et al.* Crystal structure of Spot 14, a modulator of fatty acid synthesis. *Proc. Natl Acad. Sci. USA* **107,** 18820–18825 (2010).
7. Kokoeva, M. V., Yin, H. & Flier, J. S. Neurogenesis in the hypothalamus of adult mice: potential role in energy balance. *Science* **310,** 679–683 (2005).
8. Gan, B. *et al.* Lkb1 regulates quiescence and metabolic homeostasis of haematopoietic stem cells. *Nature* **468,** 701–704 (2010).
9. Nakada, D., Saunders, T. L. & Morrison, S. J. Lkb1 regulates cell cycle and energy metabolism in haematopoietic stem cells. *Nature* **468,** 653–658 (2010).
10. Gurumurthy, S. *et al.* The Lkb1 metabolic sensor maintains haematopoietic stem cell survival. *Nature* **468,** 659–663 (2010).
11. Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Rev. Cancer* **7,** 763–777 (2007).
12. Chakravarthy, M. V. *et al.* Brain fatty acid synthase activates PPARα to maintain energy homeostasis. *J. Clin. Invest.* **117,** 2539–2552 (2007).

13. Lagace, D. C. *et al.* Dynamic contribution of nestin-expressing stem cells to adult neurogenesis. *J. Neurosci.* **27,** 12623–12629 (2007).
14. Suh, H. *et al. In vivo* fate analysis reveals the multipotent and self-renewal capacities of Sox^{2+} neural stem cells in the adult hippocampus. *Cell Stem Cell* **1,** 515–528 (2007).
15. Cordey, M., Limacher, M., Kobel, S., Taylor, V. & Lutolf, M. P. Enhancing the reliability and throughput of neurosphere culture on hydrogel microwell arrays. *Stem Cells* **26,** 2586–2594 (2008).
16. Kim, C. W. *et al.* Induced polymerization of mammalian acetyl-CoA carboxylase by MIG12 provides a tertiary level of regulation of fatty acid synthesis. *Proc. Natl Acad. Sci. USA* **107,** 9626–9631 (2010).
17. Aipoalani, D. L., O'Callaghan, B. L., Mashek, D. G., Mariash, C. N. & Towle, H. C. Overlapping roles of the glucose-responsive genes, S14 and S14R, in hepatic lipogenesis. *Endocrinology* **151,** 2071–2077 (2010).
18. Yamaguchi, M., Saito, H., Suzuki, M. & Mori, K. Visualization of neurogenesis in the central nervous system using nestin promoter–GFP transgenic mice. *Neuroreport* **11,** 1991–1996 (2000).
19. Johansson, T. *et al.* Building a zoo of mice for genetic analyses: a comprehensive protocol for the rapid generation of BAC transgenic mice. *Genesis* **48,** 264–280 (2010).
20. Karalay, O. *et al.* Prospero-related homeobox 1 gene (Prox1) is regulated by canonical Wnt signaling and has a stage-specific role in adult hippocampal neurogenesis. *Proc. Natl Acad. Sci. USA* **108,** 5807–5812 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Aigner, D. C. Lie, F. H. Gage and members of the Jessberger group for conceptual input; S. Kobel, C. Fischer, K. Walter, P. Sidiropoulos, T. Buch, B. Becher, P. Pelczar, P. Lötscher, A. J. Eisch and D. C. Lagace for experimental help or reagents; and the Light Microscopy and Screening Center (LMSC) of the ETH Zurich and the BioImaging and Optics Platform (BIOP) of the EPFL for help with imaging. This study was supported by the NCCR Neural Plasticity and Repair, Swiss National Science Foundation, TH grant (ETH-01 08-1), Zurich Neuroscience Center (ZNZ), Novartis Foundation, Theodore Ott Foundation, and the EMBO Young Investigator program (to S.J.). M.K. was supported by the Janggen-Pöhn foundation.

Author Contributions M.K. contributed to the concept, carried out experiments, analysed data and co-wrote the paper. S.M.G.B. carried out experiments and analysed data. L.Z., C.v.S. and R.A.C.M. carried out experiments. N.Z. carried out the metabolomics experiments. M.J.A.B. analysed the array data. M.R. and M.P.L. contributed to the time-lapse imaging of NSPCs. W.J.K. contributed to the lipid metabolism experiments. Ö.K., U.S. and C.F.S. provided reagents. All authors revised the manuscript. S.J. developed the concept and wrote the paper.

Author Information Microarray data are deposited in the Gene Expression Omnibus under accession number GSE27391. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J. (jessberger@hifo.uzh.ch).

doi:10.1038/nature11690

Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants

Wenqing Fu[1], Timothy D. O'Connor[1], Goo Jun[2], Hyun Min Kang[2], Goncalo Abecasis[2], Suzanne M. Leal[3], Stacey Gabriel[4], David Altshuler[4], Jay Shendure[1], Deborah A. Nickerson[1], Michael J. Bamshad[1,5], NHLBI Exome Sequencing Project* & Joshua M. Akey[1]

Establishing the age of each mutation segregating in contemporary human populations is important to fully understand our evolutionary history[1,2] and will help to facilitate the development of new approaches for disease-gene discovery[3]. Large-scale surveys of human genetic variation have reported signatures of recent explosive population growth[4–6], notable for an excess of rare genetic variants, suggesting that many mutations arose recently. To more quantitatively assess the distribution of mutation ages, we resequenced 15,336 genes in 6,515 individuals of European American and African American ancestry and inferred the age of 1,146,401 autosomal single nucleotide variants (SNVs). We estimate that approximately 73% of all protein-coding SNVs and approximately 86% of SNVs predicted to be deleterious arose in the past 5,000–10,000 years. The average age of deleterious SNVs varied significantly across molecular pathways, and disease genes contained a significantly higher proportion of recently arisen deleterious SNVs than other genes. Furthermore, European Americans had an excess of deleterious variants in essential and Mendelian disease genes compared to African Americans, consistent with weaker purifying selection due to the Out-of-Africa dispersal. Our results better delimit the historical details of human protein-coding variation, show the profound effect of recent human history on the burden of deleterious SNVs segregating in contemporary populations, and provide important practical information that can be used to prioritize variants in disease-gene discovery.

As part of the US National Institutes of Health (NIH) Heart, Lung and Blood Institute (NHLBI)-sponsored Exome Sequencing Project (ESP), we sequenced the exomes of 6,515 individuals (Supplementary Table 1), including 4,298 European Americans and 2,217 African Americans. Exome data were subjected to standard quality control filters as previously described[6] (Supplementary Information), resulting in a data set of 1,146,401 autosomal protein-coding SNVs with a known ancestral state (709,816 and 643,128 in European Americans and African Americans, respectively) distributed across 15,336 protein-coding genes. To quantitatively estimate the age of each SNV (that is, the allele age), we developed a simulation approach to generate a series of coalescent trees for a specified demographic model, and estimated allele age based on a derivation described previously[7] (Supplementary Information). We verified the accuracy and robustness of this approach to factors including recombination rate heterogeneity, population growth, migration and purifying selection. Extensive coalescent simulations showed that we could accurately estimate the expected allele age from the simulated data, although the variance associated with any individual SNV can be large (Supplementary Figs 6 and 7).

We estimated the age of all 1,146,401 SNVs using 6 different demographic models[5,6,8–11], 3 of which considered recent explosive population growth[5,6,8] (Supplementary Table 2). Estimates of allele age were generally robust across different demographic models, with the largest discrepancies resulting in a twofold difference in average age across all SNVs (Supplementary Table 3 and Supplementary Fig. 8a). However, because most SNVs arose recently (see below), differences among demographic models were highly concordant (Supplementary Information). Accordingly, we report results based on a modified Out-of-Africa model[9] in which accelerated population growth began 5,115 years ago with a per-generation growth rate of 1.95% and 1.66% for European Americans and African Americans, respectively[6].

The site frequency spectrum (SFS) of protein-coding SNVs revealed an enormous excess of rare variants (Fig. 1a). Indeed, we observed an SNV approximately once every 52 base pairs (bp) and 57 bp in European Americans and African Americans, respectively, whereas in a population without recent explosive growth we would expect the SNVs to occur once every 257 bp and 152 bp in European Americans and African Americans, respectively (Supplementary Information). Thus, the European American and African American samples contain approximately fivefold and threefold increases in SNVs, respectively, attributable to explosive population growth, resulting in a large burden of rare SNVs predicted to have arisen very recently (Fig. 1b). For example, the expected age of derived singletons, which comprise 55.1% of all SNVs, is 1,244 and 2,107 years for the European American and African American samples, respectively. Overall, 73.2% of SNVs (81.4% and 58.7% in European Americans and African Americans, respectively) are predicted to have arisen in the past 5,000 years. SNVs that arose more than 50,000 years ago were observed more frequently in the African American samples (Fig. 1b), which probably reflects stronger genetic drift in European Americans associated with the Out-of-Africa dispersal.

The average age across all SNVs was 34,200 $\pm$ 900 years ($\pm$ s.d.) in European Americans and 47,600 $\pm$ 1,500 years in African Americans, and these estimates were robust to sequencing errors (Supplementary Information and Supplementary Fig. 9). As expected, SNVs shared between European Americans and African Americans were significantly older (104,400 years and 115,800 years for European Americans and African Americans, respectively) than population-specific variants (5,400 years and 15,300 years in European Americans and African Americans, respectively; Fig. 1c) (t-test; $P < 0.00001$ by permutation). Furthermore, there were large and significant differences among the average allele ages of SNVs stratified by functional type (t-test; $P < 0.00001$ by permutation). For example, splice site, nonsense and non-synonymous SNVs were two to eight times younger than synonymous and non-coding variants (Fig. 1d). Moreover, we classified amino acids into four groups (non-polar and neutral, polar and neutral, acidic and polar, and basic and polar), and non-synonymous SNVs that resulted in changes between groups were significantly younger than

[1]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [2]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA. [3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. [4]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [5]Department of Pediatrics, University of Washington, Seattle, Washington 98195, USA.
*Lists of participants and affiliations appear in the Supplementary Information.

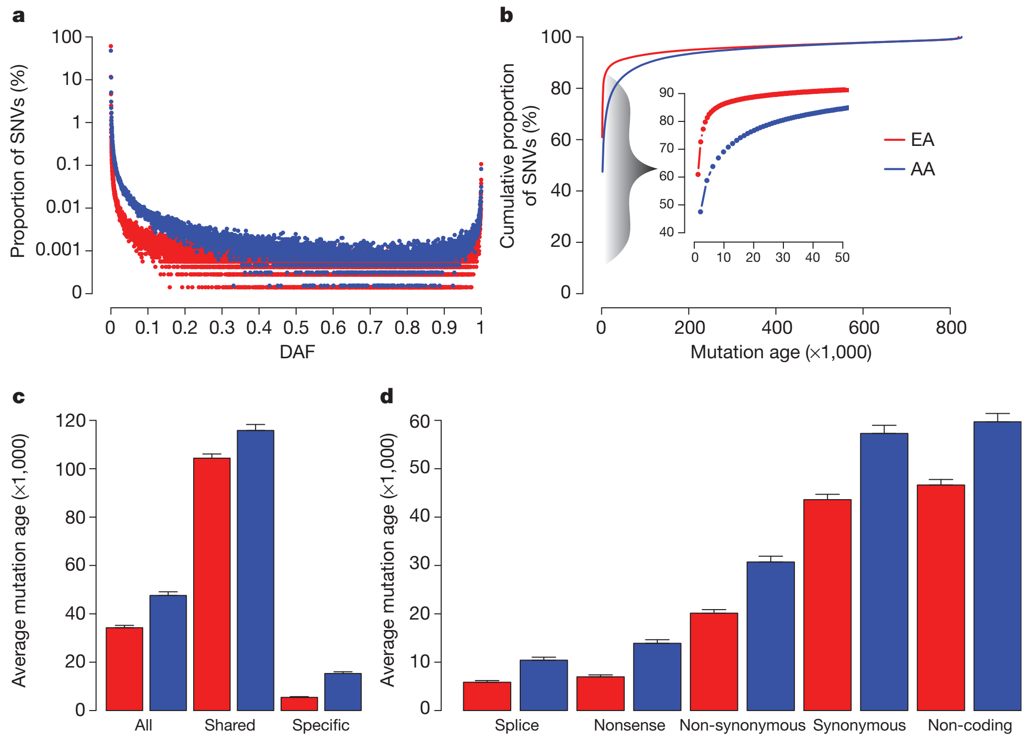


Figure 1 | The vast majority of protein-coding single-nucleotide variants arose recently. a, The site-frequency spectrum for European Americans (EAs, red) and African Americans (AAs, blue). DAF, derived allele frequency. **b**, Cumulative proportion of SNVs for a given allele age. The inset highlights the cumulative proportion of SNVs that are estimated to have arisen in the last 50,000 years. The *x* axis denotes allele age (×1,000) and the *y* axis indicates the the cumulative proportion of SNVs (%). **c**, Average age for all SNVs, SNVs found in both the European Americans and African Americans (shared), and SNVs found in only one population (specific). **d**, Average age for different types of variants. Error bars represent s.d.

those that resulted in changes within groups (*t*-test; $P < 0.00001$ by permutation; Supplementary Fig. 10a). These differences in average allele age are probably due to varying intensities of selective constraint among different classes of SNVs[12]. Consistent with this prediction, we observed significantly higher values of the neutrality index, a measure of the direction and degree of departure from neutral evolution, in genomic regions enriched for younger variants (Spearman's correlation; $P = 0.004$ and $P = 0.001$ for European Americans and African Americans, respectively; Supplementary Fig. 11), indicating a higher burden of deleterious SNVs.

To more directly identify putatively deleterious SNVs, we used four functional prediction methods (SIFT[13], PolyPhen2 (ref. 14), a likelihood ratio test[15], MutationTaster[16]) applicable to non-synonymous SNVs and two conservation-based methods (GERP++[17] and PhyloP[18]) applicable to all SNVs (Supplementary Information). We found a strong inverse relationship between average SNV age and the number of methods that predicted a variant to be deleterious (Fig. 2a, b). Thus, SNVs predicted to be deleterious by multiple methods probably experience (on average) more intense purifying selection and may be of particular interest in disease mapping studies. The age of non-synonymous SNVs predicted to be deleterious by all 6 methods was 3,000 and 6,200 years in European Americans and African Americans, respectively, and 88.7% were less than 5,000 years old (92.9% and 80.6% in European Americans and African Americans, respectively).

The strengths and weaknesses of functional prediction methods vary substantially and as a result the accuracy of any single method

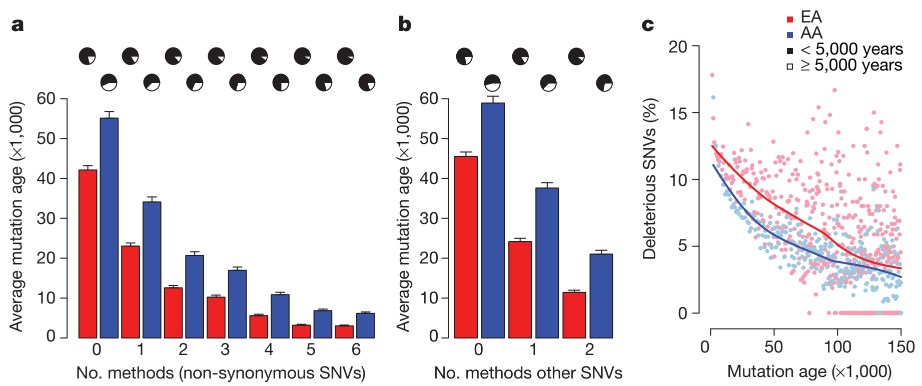


Figure 2 | Characteristics of allele age for deleterious single-nucleotide variants. a, **b**, Average age of non-synonymous and other SNVs as a function of the number of methods that predict the variant to be deleterious. Pie charts represent the proportion of SNVs that arose less than (black) or more than (white) 5,000 years ago. Error bars represent s.d. **c**, Relationship between the proportion of SNVs predicted to be deleterious and SNV age. Note, greater than 99% of deleterious SNVs are estimated to have arisen in the past 150,000 years. Solid lines represent a locally weighted scatterplot smoothing (LOESS) fit to the data.

is modest[15]. Therefore, we used a majority rule approach to identify a more conservative set of SNVs predicted to be deleterious[6]. Specifically, non-synonymous SNVs predicted to be functionally significant by at least four methods and all other SNVs (synonymous, splice and non-coding variants) predicted by two conservation-based methods were designated as deleterious. In total, 14.4% (164,688) of SNVs, including 152,633 nonsynonymous variants, met these criteria. We found that allele age was strongly related to the probability that a variant was predicted to be deleterious (Supplementary Fig. 12), with the fraction of SNVs predicted to be deleterious diminishing as allele age increased (Fig. 2c and Supplementary Fig. 13). The average age of conservatively defined deleterious variants was 5,200 ± 300 years for European Americans and 10,100 ± 600 years for African Americans. Moreover, 86.4% of these SNVs were predicted to have arisen in the past 5,000 years (91.2% and 77.0% for European Americans and African Americans, respectively), corresponding to the onset of accelerated population growth (Fig. 3a). In other demographic models, a similarly high proportion of deleterious SNVs were predicted to have arisen since the onset of accelerated growth rates, with the exact timing varying somewhat among models, but always in the timeframe of 5,000 to 10,000 years (Supplementary Table 3 and Supplementary Fig. 8b, c).

Moreover, 7,197 (57.4%) of the 12,533 genes in European Americans and 4,534 (37.5%) of the 11,607 genes in African Americans that harbour one or more deleterious variants only possess deleterious SNVs with an estimated age of less than 5,000 years (Fig. 3b). Thus, recent accelerated population growth has had a large influence on the number of genes harbouring deleterious variants in contemporary populations. Notably, after correcting for exon length of each gene, 3 and 18 genes in European Americans and in African Americans, respectively, have a significant excess of deleterious variants that arose after the onset of recent accelerated growth ($P \leq 3 \times 10^{-6}$; Supplementary Table 4), including 12 genes that have been associated with human diseases[19] such as *LAMC1* (premature ovarian failure[20]), *LRP1* (Alzheimer's disease[21]), *CPE* (coronary artery atherosclerosis[22]) and *KIAA0196* (hereditary spastic paraplegia[23]).

Next, we investigated the distribution of ages for conservatively defined deleterious SNVs in 849 genes that cause Mendelian disorders[24], 2,663 genes associated with complex diseases[19], 1,226 genes considered 'essential' (that is, a mouse knockout associated with lethality or sterility)[25] and 11,711 genes classified as 'other' (Supplementary Information). The proportions of deleterious SNVs in genes for Mendelian disorders (15.9%), essential genes (15.2%) and genes associated with complex diseases (15.1%) were each significantly higher (Fisher's exact test, $P < 10^{-16}$) than the proportion in other genes (14.0%). In the European American samples, the proportion of deleterious SNVs did not decline monotonically as a function of age for Mendelian and essential genes. Instead, the proportion of deleterious variants with an estimated age of 50,000 to 100,000 years in Mendelian disease genes and 100,000 to 150,000 years in essential genes were elevated (Fig. 4a). This pattern was not observed in the African American samples (Fig. 4a). To explore this observation, we carried out simulations to estimate the probability that a deleterious SNV survives to the present day as a function of when the variant arose, the magnitude of selection and the presence or absence of an Out-of-Africa bottleneck (Supplementary

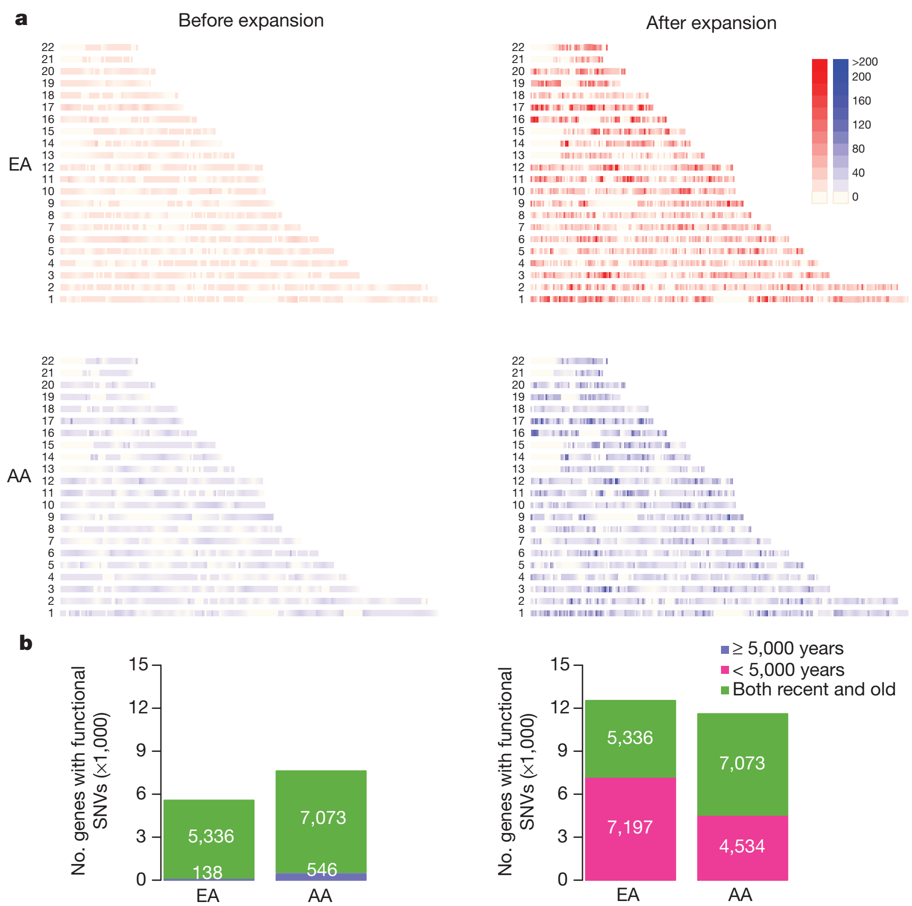


Figure 3 | Distribution of deleterious single-nucleotide variants across the exome before and after recent accelerated population growth. a, Heat map representation of deleterious SNV density (number of deleterious protein-coding variants per Mb) for European Americans (EAs, red) and African Americans (AAs, blue). The distributions of deleterious SNVs across the exome before and after recent accelerated population growth are shown in the left and right panels, respectively. Numbers on the *y* axis represent chromosome number. **b**, The bar plots summarize the number of genes segregating one or more deleterious SNVs that arose before (left) or after (right) recent accelerated population growth.

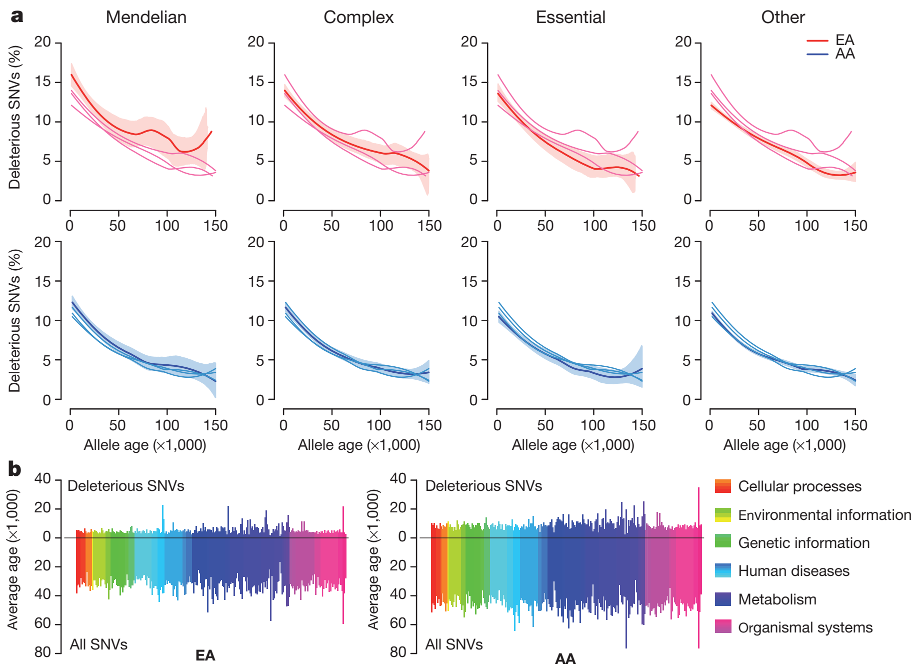


Figure 4 | Heterogeneity of allele age across genes and pathways. **a**, Distribution of the proportion of deleterious SNVs for Mendelian, complex, essential and other genes in European Americans (EAs, top) and African Americans (AAs, bottom) versus age. Data for each of the four categories of genes is shown in each plot, with darker lines representing the specific gene class indicated by the column label. Shaded regions define 95% confidence intervals obtained by bootstrapping. **b**, Average ages for deleterious (projecting up) and all (projecting down) SNVs across 235 KEGG pathways that can be organized into six broad classes. Each of the six classes is comprised of multiple sub-classes, indicated by the different colour shadings (Supplementary Information).

Information). Simulations of deleterious alleles in the presence of a bottleneck recapitulated the patterns observed in European Americans (Supplementary Fig. 14). Specifically, in the presence of a bottleneck, weakly deleterious alleles (selection coefficient, $s \leq 0.001$) have an increased probability of survival precisely in the intervals 50,000 to 100,000 years and 100,000 to 150,000 years. Thus, our simulations suggest that genes underlying disease and essential genes are more functionally constrained relative to other genes, and the bottleneck associated with the Out-of-Africa dispersal led to less efficient purging of weakly deleterious alleles[26].

Finally, we found that the average age of deleterious variants (and the proportion of deleterious variants; Supplementary Fig. 15) was significantly different across 235 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (Kruskal–Wallis rank sum test; $P = 0.0025$ and $P < 0.00001$ for European Americans and African Americans, respectively; Fig. 4b and Supplementary Information). The average age across pathways did not vary significantly when all SNVs were considered (Kruskal–Wallis rank sum test; $P = 0.259$ and $P = 0.075$ for European Americans and African Americans, respectively), indicating that the differences observed for deleterious variants probably represent heterogeneity of functional constraint across pathways. In general, the average age of deleterious variants in metabolic pathways was older than that in other pathways (Mann–Whitney test, $P \leq 0.0001$ for European Americans and African Americans, respectively), suggesting they are subject to less functional constraint. Conversely, deleterious variants in human disease pathways (Mann–Whitney test, $P = 0.03$ for African Americans) and in pathways involved in organismal systems were significantly younger than other pathways (Mann–Whitney test, $P = 0.04$ and $P = 0.002$ for European Americans and African Americans, respectively).

In summary, the spectrum of protein-coding variation is considerably different today compared to what existed as recently as 200 to 400 generations ago. Of the putatively deleterious protein-coding SNVs, 86.4% arose in the last 5,000 to 10,000 years, and they are enriched for mutations of large effect (Supplementary Fig. 14) as selection has not had sufficient time to purge them from the population. Thus, it seems likely that rare variants have an important role in heritable phenotypic variation, disease susceptibility and adverse drug responses. In principle, our results provide a framework for developing new methods to prioritize potential disease-causing variants in gene-mapping studies. More generally, the recent dramatic increase in human population size, resulting in a deluge of rare functionally important variation, has important implications for understanding and predicting current and future patterns of human disease and evolution. For example, the increased mutational capacity of recent human populations has led to a larger burden of Mendelian disorders, increased the allelic and genetic heterogeneity of traits, and may have created a new repository of recently arisen advantageous alleles that adaptive evolution will act upon in subsequent generations[27].

METHODS SUMMARY

Exome sequences were obtained for 6,823 individuals, who were sequenced to high coverage (median depth greater than 100×) on an Illumina GAII or HiSeq2000. Library construction, exome capture, sequencing, mapping, calling and filtering were carried out as described previously, with minor modifications[6] (see also Supplementary Information). After quality control and removal of related individuals, 6,515 individuals were retained. Ancestry of each individual was inferred by principal component analysis (PCA) performed on the sequence data. We developed a simulation approach based on coalescent theory to estimate allele age, which was applied to 1,146,401 autosomal SNVs with known ancestral states. A complete description of the materials and methods is provided in the Supplementary Information.

Received 13 July; accepted 19 October 2012.
Published online 28 November 2012.

1. Kimura, M. & Ota, T. The age of a neutral mutant persisting in a finite population. *Genetics* **75,** 199–212 (1973).
2. Tishkoff, S. A. & Verrelli, B. C. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu. Rev. Genomics Hum. Genet.* **4,** 293–340 (2003).
3. Slatkin, M. & Rannala, B. Estimating allele age. *Annu. Rev. Genomics Hum. Genet.* **1,** 225–249 (2000).
4. Keinan, A. & Clark, A. G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336,** 740–743 (2012).

5. Nelson, M. R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337,** 100–104 (2012).
6. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337,** 64–69 (2012).
7. Griffiths, R. C. & Tavaré, S. The age of a mutation in a general coalescent tree. *Commun. Stat. Stoch. Models* **14,** 273–295 (1998).
8. Coventry, A. *et al.* Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nature Commun.* **1,** 131 (2010).
9. Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc. Natl Acad. Sci. USA* **108,** 11983–11988 (2011).
10. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5,** e1000695 (2009).
11. Schaffner, S. F. *et al.* Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* **15,** 1576–1583 (2005).
12. Gibson, G. Rare and common variants: twenty arguments. *Nature Rev. Genet.* **13,** 135–145 (2012).
13. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4,** 1073–1081 (2009).
14. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7,** 248–249 (2010).
15. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19,** 1553–1561 (2009).
16. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods* **7,** 575–576 (2010).
17. Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLOS Comput. Biol.* **6,** e1001025 (2010).
18. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20,** 110–121 (2010).
19. Becker, K. G., Barnes, K. C., Bright, T. J. & Wang, S. A. The genetic association database. *Nature Genet.* **36,** 431–432 (2004).
20. Pyun, J. A., Cha, D. H. & Kwack, K. LAMC1 gene is associated with premature ovarian failure. *Maturitas* **71,** 402–406 (2012).
21. Liu, Q. *et al.* Amyloid precursor protein regulates brain apolipoprotein E and cholesterol metabolism through lipoprotein receptor LRP1. *Neuron* **56,** 66–78 (2007).
22. Jia, E. Z. *et al.* Association of the mutation for the human carboxypeptidase E gene exon 4 with the severity of coronary artery atherosclerosis. *Mol. Biol. Rep.* **36,** 245–254 (2009).
23. Valdmanis, P. N. *et al.* Mutations in the *KIAA0196* gene at the *SPG8* locus cause hereditary spastic paraplegia. *Am. J. Hum. Genet.* **80,** 152–161 (2007).
24. Blekhman, R. *et al.* Natural selection on genes that underlie human disease susceptibility. *Curr. Biol.* **18,** 883–889 (2008).
25. Liao, B. Y., Scott, N. M. & Zhang, J. Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. *Mol. Biol. Evol.* **23,** 2072–2080 (2006).
26. Lohmueller, K. E. *et al.* Proportionally more deleterious genetic variation in European than in African populations. *Nature* **451,** 994–997 (2008).
27. Hawks, J., Wang, E. T., Cochran, G. M., Harpending, H. C. & Moyzis, R. K. Recent acceleration of human adaptive evolution. *Proc. Natl Acad. Sci. USA* **104,** 20753–20758 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the support of the National Heart, Lung and Blood Institute (NHLBI), the contributions of the research institutions that participated in this study, the study investigators, field staff and study participants who created this resource for biomedical research, and the Population Genetics Project Team of the NHLBI. We thank J. Wilson and R. Do for critical feedback on the manuscript. Funding for the GO (Grand Opportunity) Exome Sequencing Project was provided by NHLBI grants RC2 HL-103010 (Heart GO), RC2 HL-102923 (Lung GO) and RC2 HL-102924 (WHISP). The exome sequencing was was supported by NHLBI grants RC2 HL-102925 (Broad GO) and RC2 HL-102926 (Seattle GO).

Author Contributions W.F. and J.M.A. conceived the analyses. D.A.N., S.G. and D.A. oversaw data generation and quality control. G.J., H.M.K. and G.A. developed algorithms and identified SNVs from the sequencing data. W.F. carried out the majority of analyses with contributions from T.D.O. W.F., M.J.B., J.S. and J.M.A. analysed the data and wrote the manuscript with contributions from all authors. W.F., T.D.O., S.M.L., J.S., D.A.N., M.J.B. and J.M.A. are members of the Seattle Grand Opportunity (GO) group and G.J., H.M.K., G.A., S.G. and D.A. are members of the Broad GO group, which are both sub-groups of the NHLBI Exome Sequencing Project (ESP).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.F. (wqfu@u.washington.edu) or J.M.A. (akeyj@u.washington.edu).

Filtered sets of annotated variants and their allele frequencies are available at http://evs.gs.washington.edu/EVS/ and genotypes and phenotypes from a large subset of individuals are also available through dbGaP (http://www.ncbi.nlm.nih.gov/gap) using the following accession information: NHLBI GO-ESP: Women's Health Initiative Exome Sequencing Project (WHI) – WHISP, WHISP_Subject_Phenotypes, pht002246.v2.p2, phs000281.v2.p2; NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (JHS), ESP_HeartGO_JHS_LDLandEOMI_Subject_Phenotypes, pht002539.v1.p1, phs000402.v1.p1; NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (FHS), HeartGO_FHS_LDLandEOMI_PhenotypeDataFile, pht002476.v1.p1, phs000401.v1.p1; NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (CHS), HeartGO_CHS_LDL_PhenotypeDataFile, pht002536.v1.p1, phs000400.v1.p1; NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (ARIC), ESP_ARIC_LDLandEOMI_Sample, pht002466.v1.p1, phs000398.v1.p1; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Cystic Fibrosis), ESP_LungGO_CF_PA_Culture_Data, pht002227.v1.p1, phs000254.v1.p1; NHLBI GO-ESP: Early-Onset Myocardial Infarction (Broad EOMI), ESP_Broad_EOMI_Subject_Phenotypes, pht001437.v1.p1, phs000279.v1.p1; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Pulmonary Arterial Hypertension), PAH_Subject_Phenotypes_Baseline_Measures, pht002277.v1.p1, phs000290.v1.p1; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Lung Health Study of Chronic Obstructive Pulmonary Disease), LHS_COPD_Subject_Phenotypes_Baseline_Measures, pht002272.v1.p1, phs000291.v1.p1.

doi:10.1038/nature11697

Bright radio emission from an ultraluminous stellar-mass microquasar in M 31

Matthew J. Middleton[1,2], James C. A. Miller-Jones[3], Sera Markoff[2], Rob Fender[4], Martin Henze[5], Natasha Hurley-Walker[3], Anna M. M. Scaife[4], Timothy P. Roberts[1], Dominic Walton[6,7], John Carpenter[7], Jean-Pierre Macquart[3,8], Geoffrey C. Bower[9], Mark Gurwell[10], Wolfgang Pietsch[5], Frank Haberl[5], Jonathan Harris[1], Michael Daniel[1], Junayd Miah[1], Chris Done[1], John S. Morgan[3], Hugh Dickinson[11], Phil Charles[4,12], Vadim Burwitz[5], Massimo Della Valle[13,14], Michael Freyberg[5], Jochen Greiner[5], Margarita Hernanz[15], Dieter H. Hartmann[16], Despina Hatzidimitriou[17], Arno Riffeser[18], Gloria Sala[19], Stella Seitz[18], Pablo Reig[20], Arne Rau[5], Marina Orio[21], David Titterington[22] & Keith Grainge[22]

A subset of ultraluminous X-ray sources (those with luminosities of less than 10^{40} erg s^{-1}; ref. 1) are thought to be powered by the accretion of gas onto black holes with masses of ~5–20$M_{\odot}$, probably by means of an accretion disk[2,3]. The X-ray and radio emission are coupled in such Galactic sources; the radio emission originates in a relativistic jet thought to be launched from the innermost regions near the black hole[4,5], with the most powerful emission occurring when the rate of infalling matter approaches a theoretical maximum (the Eddington limit). Only four such maximal sources are known in the Milky Way[6], and the absorption of soft X-rays in the interstellar medium hinders the determination of the causal sequence of events that leads to the ejection of the jet. Here we report radio and X-ray observations of a bright new X-ray source in the nearby galaxy M 31, whose peak luminosity exceeded 10^{39} erg s^{-1}. The radio luminosity is extremely high and shows variability on a timescale of tens of minutes, arguing that the source is highly compact and powered by accretion close to the Eddington limit onto a black hole of stellar mass. Continued radio and X-ray monitoring of such sources should reveal the causal relationship between the accretion flow and the powerful jet emission.

XMM-Newton first detected XMMU J004243.6+412519 on 15 January 2012 (ref. 7) at an X-ray luminosity of 2×10^{38} erg s^{-1} (for a distance to M 31 of 0.78 Mpc (ref. 8)), with an X-ray spectrum that could be fully described by a hard power-law, characteristic of sub-Eddington accretion (mass accretion rates <70% of the Eddington limit[9,10]). The source then rose to $>10^{39}$ erg s^{-1} in two subsequent detections, fulfilling the traditional definition of an ultraluminous X-ray source (ULX; although other definitions exist, the term ULX is numerical rather than physically motivated)[11] and significantly above the cutoff luminosity of the X-ray luminosity function of M 31 (ref. 12), which shows no sources more luminous than 2×10^{38} erg s^{-1}.

At the peak luminosity of $(1.26 \pm 0.01) \times 10^{39}$ erg s^{-1}, the X-ray spectrum seemed similar to that of Galactic black-hole X-ray binaries (BHXRBs) at mass accretion rates close to or above the Eddington limit[13]. In such cases, the emission is dominated by an optically thick accretion disk[2] whose spectrum may appear broadened because the accretion process is no longer radiatively efficient[14]. This can be accompanied by a second, weaker, thermal component at higher energies, possibly due to Compton up-scattering of disk photons in a wind or photosphere. These two components are also required to fit high-quality spectra of nearby 'low-luminosity' ULXs[3,15], implying that similar accretion processes are taking place. Although the intrinsic emission can potentially be amplified through geometrical beaming[16], this is thought to be important only for ULXs above 10^{40} erg s^{-1}.

After the X-ray outburst, the source was monitored by the Swift and Chandra missions for more than 8 weeks (see Fig. 1 and Supplementary Table 1), until the source was no longer observable because of the Sun angle. The X-ray luminosity decreased slightly to $\sim 10^{39}$ erg s^{-1}, with the spectrum evolving to be fully described by emission from a standard accretion disk[2] with a peak temperature characteristic of high-mass-accretion-rate Galactic BHXRBs[9]. Both the timescale for the X-ray spectral variations and the temperature of the disk component rule out an interpretation as a background active galaxy.

A subsequent period of X-ray monitoring showed the source to remain disk-dominated while decaying to a luminosity of $\sim 7 \times 10^{37}$ erg s^{-1}. The correlation of luminosity with disk temperature as $L \propto T^4$ is a well-established observational tracer of sub-Eddington accretion. We find that the best-fitting spectral parameters (see Supplementary Table 2) deviated strongly from this correlation when the source was brightest: the temperature decreased with luminosity (Supplementary Fig. 2). Such behaviour is well documented in ULXs[17], strengthening our identification of this object as a member of this class, albeit a low-luminosity one.

The X-ray detection of this nearby bright source motivated a series of radio observations by the Karl G. Jansky Very Large Array (VLA; see Supplementary Table 3), Very Long Baseline Array (VLBA) and Arcminute Microkelvin Imager Large Array (AMI-LA). The radio light curves are shown in Fig. 2 (observational details are provided in Supplementary Information). The source was initially detected by the VLA at a highly significant level ($>40\sigma$), with a 4–5σ AMI-LA detection the following day. The spectral index (defined by $S_{\nu} \propto \nu^{\alpha}$) measured by the VLA was slightly inverted, at $\alpha = 0.27 \pm 0.16$.

The AMI-LA monitoring demonstrated that the radio emission was variable on timescales of days, as confirmed by a detection of strongly

[1]Physics Department, University of Durham, Durham DH1 3LE, UK. [2]Astronomical Institute Anton Pannekoek, Science Park 904, 1098 XH Amsterdam, The Netherlands. [3]International Centre for Radio Astronomy Research, Curtin University, GPO Box U1987, Perth, WA 6845, Australia. [4]School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK. [5]Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, 85748 Garching, Germany. [6]Institute of Astronomy, Cambridge University, Madingley Road, Cambridge CB3 0HA, UK. [7]Astronomy Department, California Institute of Technology, MC 249-17, 1200 East California Boulevard, Pasadena, California 91125, USA. [8]ARC Centre of Excellence for All-Sky Astrophysics (CAASTRO), Curtin University, GPO Box U1987, Perth, WA 6845, Australia. [9]Astronomy Department, B-20 Hearst Field, Annex no. 3411, University of California at Berkeley, Berkeley, California 94720-3411, USA. [10]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [11]Stockholm University, Oskar Klein Centre, AlbaNova, SE-106 91, Stockholm, Sweden. [12]Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, Republic of South Africa. [13]Osservatorio Astronomico di Capodimonte, INAF, Salita Moiariello 16, 80131 Napoli, Italy. [14]International Centre for Relativistic Astrophysics, Piazzale della Repubblica 2, 65122 Pescara, Italy. [15]Institute of Space Sciences (CSIC-IEEC), Campus UAB, Facultade de Ciències, Torre C5 parell 2, 08193 Bellaterra, Barcelona, Spain. [16]Physics and Astronomy Department, 118 Kinard Laboratory, Clemson University, Clemson, South Carolina 29631-0978, USA. [17]Department of Astrophysics, Astronomy and Mechanics, Faculty of Physics, National and Kapodistrian University of Athens, Panepistimiopolis, 157 84 Zografou, Athens, Greece. [18]University Observatory Munich, Ludwig-Maximilians-Universität, Scheinerstrasse 1, 81679 München, Germany. [19]Department of Physics and Nuclear Engineering, EUETIB (UPC-IEEC), calle Comte d'Urgell 187, 08036 Barcelona, Spain. [20]Foundation for Research and Technology-Hellas, Nikolaou Plastira 100, Vassilika Vouton, 71110 Heraklion, Crete, Greece. [21]Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, 35122 Padova, Italy. [22]Astrophysics Group, Cavendish Laboratory, Cambridge University, JJ Thomson Avenue, Cambridge CB3 0HE, UK.

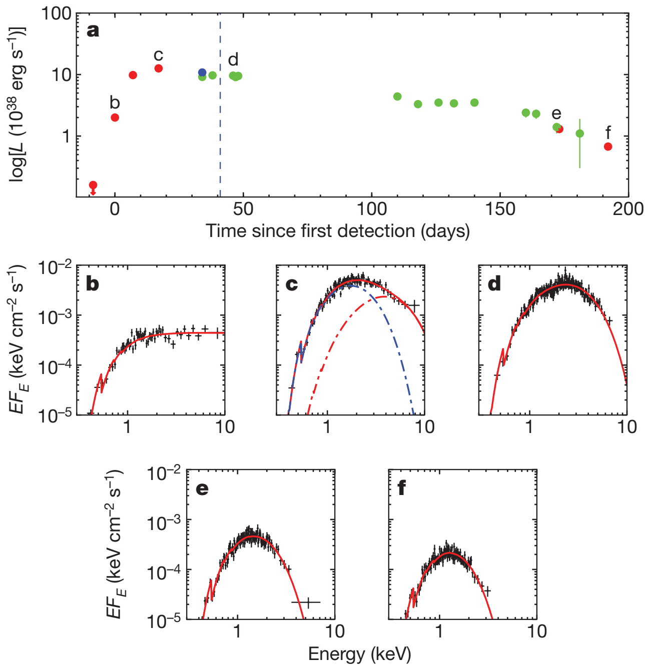


Figure 1 | Evolution of the X-ray luminosity and spectrum over time. **a**, Evolution of the source brightness over ~200 days together with the time of the first VLA detection (blue vertical dashed line). The unabsorbed 0.3-10 keV luminosities (and 90% error bars) are derived from the best-fitting models to the observations with XMM-Newton (red), Chandra (blue) and Swift (green) (see Supplementary Tables 1 and 2). An upper limit from an XMM-Newton observation taken a week before the source brightened is also included. **b–f**, Spectral data (black points with 90% error bars) and best-fitting models (in red; see Supplementary Information and Supplementary Table 2) from which the unabsorbed luminosities are calculated. **b**, The first detection before the peak brightness shows a spectrum that can be fully described by a hard power-law. **c**, As the source brightens to a peak luminosity of 1.26×10^{39} erg s^{-1}, the spectrum evolves to one requiring emission from an optically thick accretion disk (dot–dashed blue line) and a second, weaker thermal component (dot–dashed red line) that may be related to a wind or photosphere. Such spectra are also seen in high-mass-accretion-rate BHXRBs[13] and low-luminosity ULXs[3,15]. **d**, As the source dims to $\sim 10^{39}$ erg s^{-1}, the spectrum becomes fully dominated by the disk component. **e**, The spectrum remains disk-dominated but the luminosity decreases, similar to BHXRBs decaying after an outburst. **f**, The most recent XMM-Newton observation demonstrates that the source remains disk-dominated down to $\sim 7 \times 10^{37}$ erg s^{-1}, which places an upper limit on the mass[10] of $<17 M_{\odot}$.

reduced emission by the VLA (Fig. 2) and a non-detection by the Combined Array for Research in Millimeter-wave Astronomy (CARMA). Previous, well-characterized, radio-luminous ULXs have been identified with emission from an optically thin, diffuse nebula believed to be powered by the bright accretion flow[18]. The radio variability from XMMU J004243.6+412519 demonstrates that the emission cannot be nebular in origin. Furthermore, the unresolved VLBA detection shown in Fig. 2 constrains the size of the emitting region to be <1,500 AU. This size limit, together with the variability on a timescale of days, confirms that the radio emission must originate in a relatively compact, relativistically outflowing region with a brightness temperature $>6 \times 10^6$ K.

BHXRBs accreting at sub-Eddington rates produce powerful, persistent radio jets; the radio and X-ray emission are strongly correlated and are linked to black-hole mass through the 'fundamental plane of black hole activity'[19,20]. However, the thermal, disk-dominated X-ray spectrum after the peak of the outburst implies accretion at a relatively high fraction of the Eddington limit[10] (3–100%). The fundamental plane relation does not hold for sources accreting at such high mass accretion rates and so cannot be used to constrain the mass of the black hole. At these high accretion rates, the radio emission from BHXRBs is instead observed to undergo flaring periods, associated with transitions from hard to soft X-ray spectral states[5]. Indeed, scaling the radio and X-ray flux from this source to be within our Galaxy would give fluxes similar to soft, bright, radio-flaring BHXRBs such as GRS 1915+105 and Cygnus X-3 (ref. 21). The observed X-ray and radio emission of this source is therefore fully consistent with analogous behaviour, and the radio emission is probably associated with shock acceleration of particles within ejections moving away from the black hole[22] with a high Lorentz factor ($\Gamma \geq 2$; ref. 5).

The AMI-LA light curve (Fig. 2) suggests multiple ejection events, although the observed inverted, optically thick, spectrum is at odds with the optically thin emission expected from the later stages of an expanding synchrotron-emitting plasma. Either the VLA observed the early, self-absorbed stage of a flare or there is additional, free-free absorbing material in the environment of the source.

A search for shorter-timescale fluctuations in the VLA data revealed significant variability in the first epoch, on a characteristic timescale of several minutes (Fig. 2). Regardless of whether we attribute this to intrinsic variability or to scintillation (see Supplementary Information), this timescale implies a source size of only ~5 AU (a few microarcseconds at the distance of M 31), constraining the emitting region to be highly compact.

There are clear similarities between the behaviour of XMMU J004243.6+412519 and the few 'super-Eddington' Galactic BHXRBs. Of these, the canonical microquasar GRS 1915+105 is the only one whose luminosity regularly exceeds 10^{39} erg s^{-1} and that would sometimes appear as a ULX to an extragalactic observer[23]. Multiwavelength studies of the disk–jet coupling in GRS 1915+105 (ref. 24) have identified both 'plateau' states with steady jet emission at lower mass accretion rates and 'flaring' states with rapid radio oscillations at accretion rates around the Eddington limit, accompanied by an extremely soft X-ray spectrum. The latter, minute-timescale[25] radio flaring, has been observed to occur immediately after a major radio flare, accompanied by X-ray brightness changes that are dominated by variability above 3 keV. If the short-timescale radio variability we observe is intrinsic (see Supplementary Information), it could be an analogue of this behaviour, although we would not expect to detect the X-ray variability in our observations because the flux above 3 keV is extremely low as a result of the greater source distance. Although GRS 1915+105 shows clearly analogous behaviour, when scaled to a similar distance our source is approximately an order of magnitude more luminous in the radio band, with a variability brightness temperature for the flares of $\sim 7 \times 10^{10}$ K (see Supplementary Information). Such a discrepancy can readily be explained by differences in the inclination angle of the jets to our line of sight; GRS1915+105 is highly inclined and thus Doppler-deboosted, whereas we may be looking closer to the jet axis in XMMU J004243.6+412519. This physical picture is supported by comparison with the Galactic BHXRBs, Cygnus X-3 and V4641 Sgr, both of which have reached similarly high radio luminosities and probably have low angles to the line of sight[26,27]. Our source seems particularly well matched to V4641 Sgr, which reached ~0.5 Jy in its 1999 outburst[28], and has also shown similarly high brightness temperatures.

The coupled radio and X-ray behaviour in this source is fully consistent with a picture of a BHXRB rising above the Eddington limit in outburst then dimming to a bright sub-Eddington state, with radio flaring analogous to that seen in GRS 1915+105 and other BHXRBs[5]. We can therefore make a secure identification of the compact object in this source as a black hole of stellar mass ($<70 M_{\odot}$).

Because the source remains in a disk-dominated state down to $\sim 7 \times 10^{37}$ erg s^{-1}, and the lower limit for such a state is ~3% of the Eddington limit[10], this constrains the mass of the black hole to $<17 M_{\odot}$. However, because the joint radio and X-ray behaviour implies Eddington-rate accretion[5] at the peak luminosity of $\sim 1.3 \times 10^{39}$ erg s^{-1}, we favour a more conservative mass estimate of $\sim 10 M_{\odot}$.

We can also constrain the nature of the companion star from archival optical data[29]. The field does not contain a source within 3 arcsec (where the positional accuracy from the VLBA observations is extremely high;

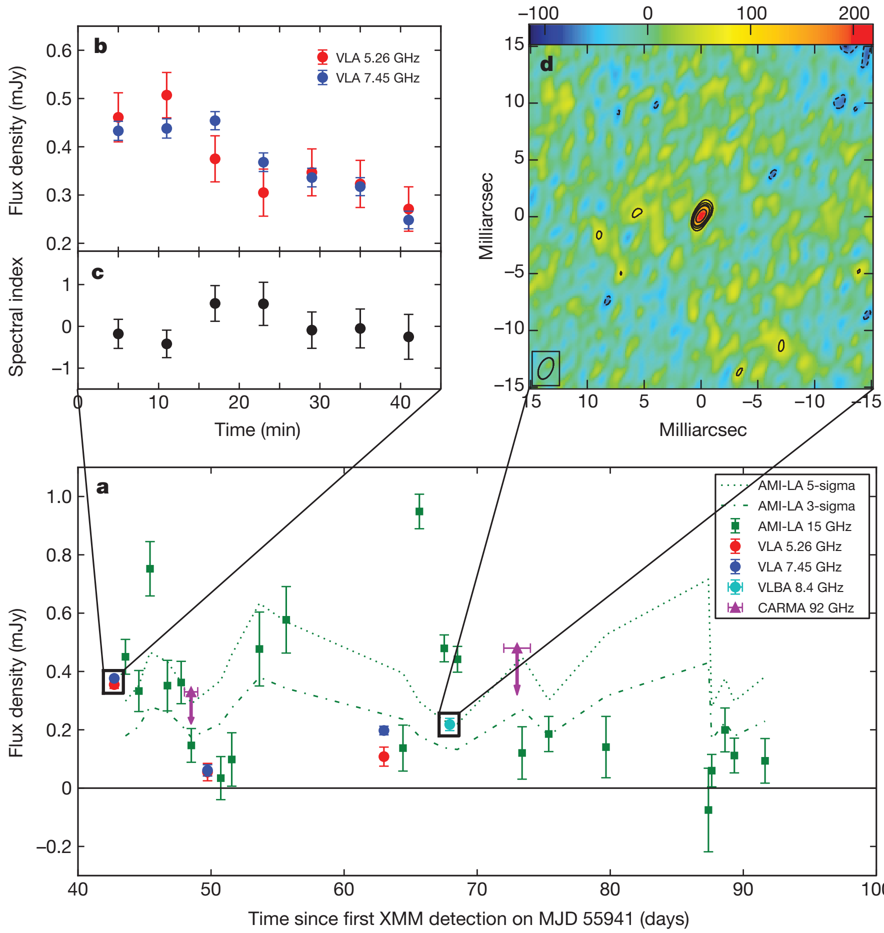


Figure 2 | Summary of the radio observations. a, Multifrequency radio light curves of the ULX from our AMI-LA, VLA, VLBA and CARMA data. For the AMI-LA data we plot the flux density from the pixel value at the source location, because of dynamic range limitations from the bright nearby confusing source, 5C 3.111 (Supplementary Fig. 1). The resulting high noise level close to the ULX hindered the standard image-plane source fitting methods used for the VLA and VLBA data. To estimate the significance of the AMI points, we have plotted both the 3-sigma and 5-sigma noise levels (dot–dashed and dotted lines, respectively), showing both a standard and a (favoured) more conservative detection threshold. **b**, Radio light curves binned to an interval of 6 min (the phase referencing cycle time), showing the short-timescale variability during the first VLA observation. There is an additional 1% systematic uncertainty on the overall flux scale. We are unable to determine conclusively whether the observed variability is intrinsic to the source or whether it arises from scintillation in the interstellar medium. Regardless of the mechanism involved, this constrains the source size to $\leq$6 μas. **c**, Corresponding spectral indices derived from the two VLA observing bands centred at 5.256 and 7.45 GHz. **d**, VLBA image of the source. Colour scale is in μJy per beam, and the contours are set at levels of $\pm(\sqrt{2})^n$ times the root mean squared noise of 21 μJy per beam, where $n = 3,4,5$. The source is unresolved down to a beam size of $2.0 \times 1.2\,\mathrm{mas}^2$ in position angle $-26°$ east of north. The (J2000) source position, measured relative to J0038+4137 and before performing any self-calibration, is RA 00 h 42 min 43.674360(9) s, dec. 41° 25′ 18.7037(1)″. Error bars in all panels show the statistical uncertainties at the 1σ level.

see Supplementary Information) down to $V \approx 25$ and $B \approx 26$, ruling out an identification of the companion as an OB star and implying that the system is accreting from a star of lower mass, similar to other transient, low-luminosity ULXs[3,30]. This also provides further evidence that the emission is not associated with a background active galaxy, which would be substantially brighter.

Because the X-ray properties of the source are typical of the class[3,15,17], we confidently extend the identification of Eddington-rate accretion onto a black hole of stellar mass in XMMU J004243.6+412519, to the larger population of low-luminosity ULXs (a substantial component of the overall population, with up to 80% having luminosities of $(1\text{–}5) \times 10^{39}\,\mathrm{erg\,s^{-1}}$)[1]. As a result, this implies that we have a large population of sources from which to develop models for Eddington accretion. In our own Galaxy we are limited to a very small number of such sources, and the large absorbing column through the plane of our Galaxy prevents a clear view of the Eddington-limited accretion flow. However, this is not the case for the vast majority of extragalactic sources, allowing the properties of the infalling matter to be accurately determined. Understanding how the disk and associated outflows behave in this regime will allow us to address outstanding cosmological problems including how outflows from quasars redistributed matter and energy in the early Universe.

Assuming that the properties of this source are representative of the larger population, we can consider the possibility of detecting similar events in future. The sensitivity of the VLA would allow us to detect a similarly relativistically beamed event out to 4 Mpc, or 0.5 Mpc if unbeamed. The X-ray monitoring cadence for M 31 has yielded two candidates in 10 years; although the monitoring was not constant, this is roughly consistent with the observed rate of Galactic BHXRBs with low-mass companions whose outbursts reach the Eddington luminosity. It is therefore a realistic prospect (see Supplementary Information) that future observations of transient ULX systems in nearby galaxies, using sensitive radio telescopes, will permit detailed disk–jet coupling studies and in doing so will significantly expand our understanding of Eddington accretion and associated phenomena.

Received 10 May; accepted 22 October 2012.
Published online 12 December 2012.

1. Walton, D., Roberts, T., Mateos, S. & Heard, V. 2XMM ultraluminous X-ray source candidates in nearby galaxies. *Mon. Not. R. Astron. Soc.* **416,** 1844–1861 (2011).

2. Shakura, N. & Sunyaev, R. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24,** 337–355 (1973).
3. Middleton, M., Sutton, A., Roberts, T., Jackson, F. & Done, C. The missing link: a low-mass X-ray binary in M 31 seen as an ultraluminous X-ray source. *Mon. Not. R. Astron. Soc.* **420,** 2969–2977 (2012).
4. Mirabel, I. & Rodriguez, L. A superluminal source in the Galaxy. *Nature* **371,** 46–48 (1994).
5. Fender, R., Belloni, T. & Gallo, E. Towards a unified model for black hole X-ray binary jets. *Mon. Not. R. Astron. Soc.* **355,** 1105–1118 (2004).
6. Grimm, H., Gilfanov, M. & Sunyaev, R. The Milky Way in X-rays for an outside observer. Log(*N*)–log(*S*) and luminosity function of X-ray binaries from RXTE/ASM data. *Astron. Astrophys.* **391,** 923–944 (2002).
7. Henze, M. *et al.* XMMU J004243.6+412519—a new X-ray transient in M 31 seen with XMM-Newton. *Astronomer's Telegram* ATel #3890, http://www.astronomerstelegram.org/?read=3890 (2012).
8. Monachesi, A. *et al.* The deepest *Hubble Space Telescope* color–magnitude diagram of M32. Evidence for intermediate-age populations. *Astrophys. J.* **727,** 55 (2011).
9. Remillard, R. & McClintock, J. X-ray properties of black-hole binaries. *Annu. Rev. Astron. Astrophys.* **44,** 49–92 (2006).
10. Dunn, R., Fender, R., Körding, E., Belloni, T. & Cabanac, C. A global spectral study of black hole X-ray binaries. *Mon. Not. R. Astron. Soc.* **403,** 61–82 (2010).
11. Feng, H. & Soria, R. Ultraluminous X-ray sources in the Chandra and XMM-Newton era. *N. Astron. Rev.* **55,** 166–183 (2011).
12. Kong, A., DiStefano, R., Garcia, M. & Greiner, J. Chandra studies of the X-ray point source luminosity functions of M 31. *Astrophys. J.* **585,** 298–304 (2003).
13. Ueda, Y., Yamaoka, K. & Remillard, R. GRS 1915+105 in 'soft state': nature of accretion disk wind and origin of X-ray emission. *Astrophys. J.* **695,** 888–899 (2009).
14. Abramowicz, M., Czerny, B., Lasota, J.-P. & Szuszkiewicz, E. Slim accretion disks. *Astrophys. J.* **332,** 646–658 (1988).
15. Middleton, M., Sutton, A. & Roberts, T. X-ray spectral evolution in the ultraluminous X-ray source M33 X-8. *Mon. Not. R. Astron. Soc.* **417,** 464–471 (2011).
16. King, A. Masses, beaming and Eddington ratios in ultraluminous X-ray sources. *Mon. Not. R. Astron. Soc.* **393,** L41–L44 (2009).
17. Feng, H. & Kaaret, P. Spectral states and evolution of ultraluminous X-ray sources. *Astrophys. J.* **696,** 1712–1726 (2009).
18. Cseh, D. *et al.* Black hole powered nebulae and a case study of the ultraluminous X-ray source IC 342 X-1. *Astrophys. J.* **749,** 17 (2012).
19. Merloni, A., Heinz, S. & di Matteo, T. A fundamental plane of black hole activity. *Mon. Not. R. Astron. Soc.* **345,** 1057–1076 (2003).
20. Plotkin, R., Markoff, S., Kelly, B., Körding, E. & Anderson, S. Using the fundamental plane of black hole activity to distinguish X-ray processes from weakly accreting black holes. *Mon. Not. R. Astron. Soc.* **419,** 267–286 (2012).
21. Gallo, E., Fender, R. & Pooley, G. A universal radio-X-ray correlation in low/hard state black hole binaries. *Mon. Not. R. Astron. Soc.* **344,** 60–72 (2003).
22. Kaiser, C., Sunyaev, R. & Spruit, H. Internal shock model for microquasars. *Astron. Astrophys.* **356,** 975–988 (2000).
23. King, A. The brightest black holes. *Mon. Not. R. Astron. Soc.* **335,** L13–L16 (2002).
24. Fender, R. & Belloni, T. GRS 1915+105 and the disc–jet coupling in accreting black hole systems. *Annu. Rev. Astron. Astrophys.* **42,** 317–364 (2004).
25. Pooley, G. & Fender, R. The variable radio emission from GRS 1915+105. *Mon. Not. R. Astron. Soc.* **292,** 925–933 (1997).
26. Corbel, S. *et al.* A giant radio flare from Cygnus X-3 with associated γ-ray emission. *Mon. Not. R. Astron. Soc.* **421,** 2947–2955 (2012).
27. Orosz, J. A. *et al.* A black hole in the superluminal source SAX J1819.3−2525 (V4641 Sgr). *Astrophys. J.* **555,** 489–503 (2001).
28. Hjellming, R. M. *et al.* Light curves and radio structure of the 1999 September transient event in V4641 Sagittarii (=XTE J1819–254=SAX J1819.3–2525). *Astrophys. J.* **544,** 977–992 (2000).
29. Vilardell, F., Ribas, I. & Jordi, C. Eclipsing binaries suitable for distance determination in the Andromeda galaxy. *Astron. Astrophys.* **459,** 321–331 (2006).
30. Soria, R. *et al.* The birth of an ultraluminous X-ray source in M83. *Astrophys. J.* **750,** 152 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Trott and R. Soria for discussions, and C. Gough for making his code available. This work was supported by a Science and Technology Facilities Council (STFC) standard grant (M.J.M.), Netherlands Organization for Scientific Research Vidi Fellowship (S.M.), European Research Council partial funding (R.F.) and grant number BMWI/DLR, FKZ 50 OR 1010 (M. Henze). The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. We thank the staff of the Mullard Radio Astronomy Observatory for their assistance in the commissioning and operation of AMI, which is supported by Cambridge University and the STFC. This work is based on observations obtained with XMM-Newton, an ESA science mission with instruments and contributions directly funded by ESA Member States and NASA. This research has also made use of data obtained from NASA's Swift and Chandra satellites.

Author Contributions M.J.M. wrote the manuscript with comments from all authors. J.C.A.M.-J. designed and analysed the VLA and VLBA observations. N.H.-W. and A.M.M.S. analysed the AMI-LA observations. J.-P.M. and J.C.A.M.-J. conducted the scintillation analysis. S.M., R.F. and M. Henze made significant contributions to the overall science case and manuscript. J.C., G.C.B. and M.G. provided support and analysis for the CARMA observations. The remaining authors either assisted with various aspects of the science case or are contributing members of the M 31 group.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.M. (m.j.middleton@durham.ac.uk).

doi:10.1038/nature11710

Topological colloids

Bohdan Senyuk[1,2], Qingkun Liu[1,3], Sailing He[3], Randall D. Kamien[4], Robert B. Kusner[5], Tom C. Lubensky[4] & Ivan I. Smalyukh[1,2,6]

Smoke, fog, jelly, paints, milk and shaving cream are common everyday examples of colloids[1], a type of soft matter consisting of tiny particles dispersed in chemically distinct host media. Being abundant in nature, colloids also find increasingly important applications in science and technology, ranging from direct probing of kinetics in crystals and glasses[2] to fabrication of third-generation quantum-dot solar cells[3]. Because naturally occurring colloids have a shape that is typically determined by minimization of interfacial tension (for example, during phase separation) or faceted crystal growth[1], their surfaces tend to have minimum-area spherical or topologically equivalent shapes such as prisms and irregular grains (all continuously deformable—homeomorphic—to spheres). Although toroidal DNA condensates and vesicles with different numbers of handles can exist[4–7] and soft matter defects can be shaped as rings[8] and knots[9], the role of particle topology in colloidal systems remains unexplored. Here we fabricate and study colloidal particles with different numbers of handles and genus g ranging from 1 to 5. When introduced into a nematic liquid crystal—a fluid made of rod-like molecules that spontaneously align along the so-called 'director'[10]—these particles induce three-dimensional director fields and topological defects dictated by colloidal topology. Whereas electric fields, photothermal melting and laser tweezing cause transformations between configurations of particle-induced structures, three-dimensional nonlinear optical imaging reveals that topological charge is conserved and that the total charge of particle-induced defects always obeys predictions of the Gauss–Bonnet and Poincaré–Hopf index theorems[11–13]. This allows us to establish and experimentally test the procedure for assignment and summation of topological charges in three-dimensional director fields. Our findings lay the groundwork for new applications of colloids and liquid crystals that range from topological memory devices[14], through new types of self-assembly[15–23], to the experimental study of low-dimensional topology[6,7,11–13].

Although a coffee mug and a doughnut look different to most of us, they are topologically equivalent solid tori or handlebodies of genus $g = 1$, both being different from, say, balls and solid cylinders of genus $g = 0$, to which they cannot be smoothly morphed without cutting[11,12]. In a similar way, molecules can form topologically distinct structures including rings, knots and other molecular configurations satisfying the constraints imposed by chemical bonds[24]. Although the topology of shapes, fields and defects is important in many phenomena and in theories ranging from the nature of elementary particles to early-Universe cosmology[25,26], topological aspects of colloidal systems (composed of particles larger than molecules and atoms but much smaller than the objects that we encounter in our everyday life) are rarely explored. Typically dealing with particles with surfaces homeomorphic to spheres, recent studies[8,9,15,18,20–23] demonstrate that the topology of curved surfaces dictates the formation of defects during two-dimensional colloidal crystallization at fluid interfaces as well as inside liquid crystal droplets and around spherical inclusions in liquid crystals. However, despite the fact that several techniques for scalable fabrication of particles with complex geometric shapes and $g > 0$ have recently been introduced[18,19,27–29], the potential impact of particle topology on colloidal alignment, self-assembly and response to fields remains unexplored.

To study the interplay of particle topology and defects in liquid crystals, we fabricated topologically distinct silica particles with planar symmetry and handlebody topology of genus g varying from 1 to 5; their surfaces had an Euler characteristic $\chi = 2 - 2g$ ranging from 0 to -8 (Fig. 1 and Supplementary Fig. 1). These particles had $1\,\mu m \times 1\,\mu m$ rounded square cross-sections and ring diameters ranging from 5 to $10\,\mu m$. Handlebody particles were introduced into a nematic liquid crystal, pentyl cyanobiphenyl, and the ensuing dispersion was infiltrated into cells bounded by parallel uniformly separated substrates treated to impose either perpendicular (homeotropic) or parallel alignment of the director $\boldsymbol{n}$ and thereby to create a uniform director $\boldsymbol{n}_0$ in their interior in the absence of inclusions. Before dispersion, the surfaces of particles were also treated to induce perpendicular boundary conditions for $\boldsymbol{n}$. The director field $\boldsymbol{n}(\boldsymbol{r})$ around these handlebody colloids, which approaches $\boldsymbol{n}_0$ at large distances, was probed optically by a combination of transmission-mode polarizing microscopy (PM) and three-photon excitation fluorescence polarizing microscopy (3PEF-PM)[30], schematically shown in Supplementary Fig. 2. Holographic optical tweezers allowed non-contact optical manipulation of particles at laser powers of 5–50 mW and local photothermal melting of the liquid crystal into an isotropic state at powers of about 100 mW and higher. Because of the strong surface anchoring and rounded cross-section of the colloids, quenching the liquid crystal from the isotropic to the nematic phase creates director configurations that vary smoothly away from homeotropic alignment at the particle surfaces and that also exhibit bulk defects.

Colloidal handlebodies spontaneously align with their ring planes either perpendicular or parallel to $\boldsymbol{n}_0$ (Figs 1 and 2 and Supplementary Figs 3–9). The prevailing alignment of handlebodies perpendicular to $\boldsymbol{n}_0$ is more common because it minimizes the elastic free energy of $\boldsymbol{n}(\boldsymbol{r})$ distortions induced by the particles with perpendicular boundary conditions. Handlebodies aligned with ring planes parallel to $\boldsymbol{n}_0$ are obtained by melting and subsequently quenching the surrounding liquid crystal with laser tweezers (Fig. 2). These particles can also be made to align parallel to or obliquely to $\boldsymbol{n}_0$ by confinement in cells of thickness comparable to their lateral dimensions (Supplementary Fig. 10). Handlebodies are elastically repelled from both confining substrates as a result of strong surface anchoring conditions. However, because of the density mismatch between silica and the liquid crystal, they tend to rest somewhat below the cell midplane, where gravity is balanced by the elastic forces (Supplementary Fig. 9).

Optical micrographs obtained by using different imaging modalities (Fig. 1a–h) reveal that handlebody colloids aligned perpendicular to $\boldsymbol{n}_0$ are all surrounded by single half-integer exterior disclination loops of topological point defect (hedgehog) charge $m = -1$ but have different defects within their interiors (Fig. 1i–l). Each genus-g particle has g defects in its holes, which are either singular disclination loops or hyperbolic point defects of topological hedgehog charge $m = +1$. Disclination loops in the holes of each handlebody can be transformed

[1]Department of Physics, University of Colorado, Boulder, Colorado 80309, USA. [2]Department of Electrical, Computer, and Energy Engineering, Materials Science and Engineering Program, and Liquid Crystal Materials Research Center, University of Colorado, Boulder, Colorado 80309, USA. [3]Centre for Optical and Electromagnetic Research, Zhejiang University, Hangzhou 310058, China. [4]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [5]Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003, USA. [6]Renewable and Sustainable Energy Institute, National Renewable Energy Laboratory and University of Colorado, Boulder, Colorado 80309, USA.

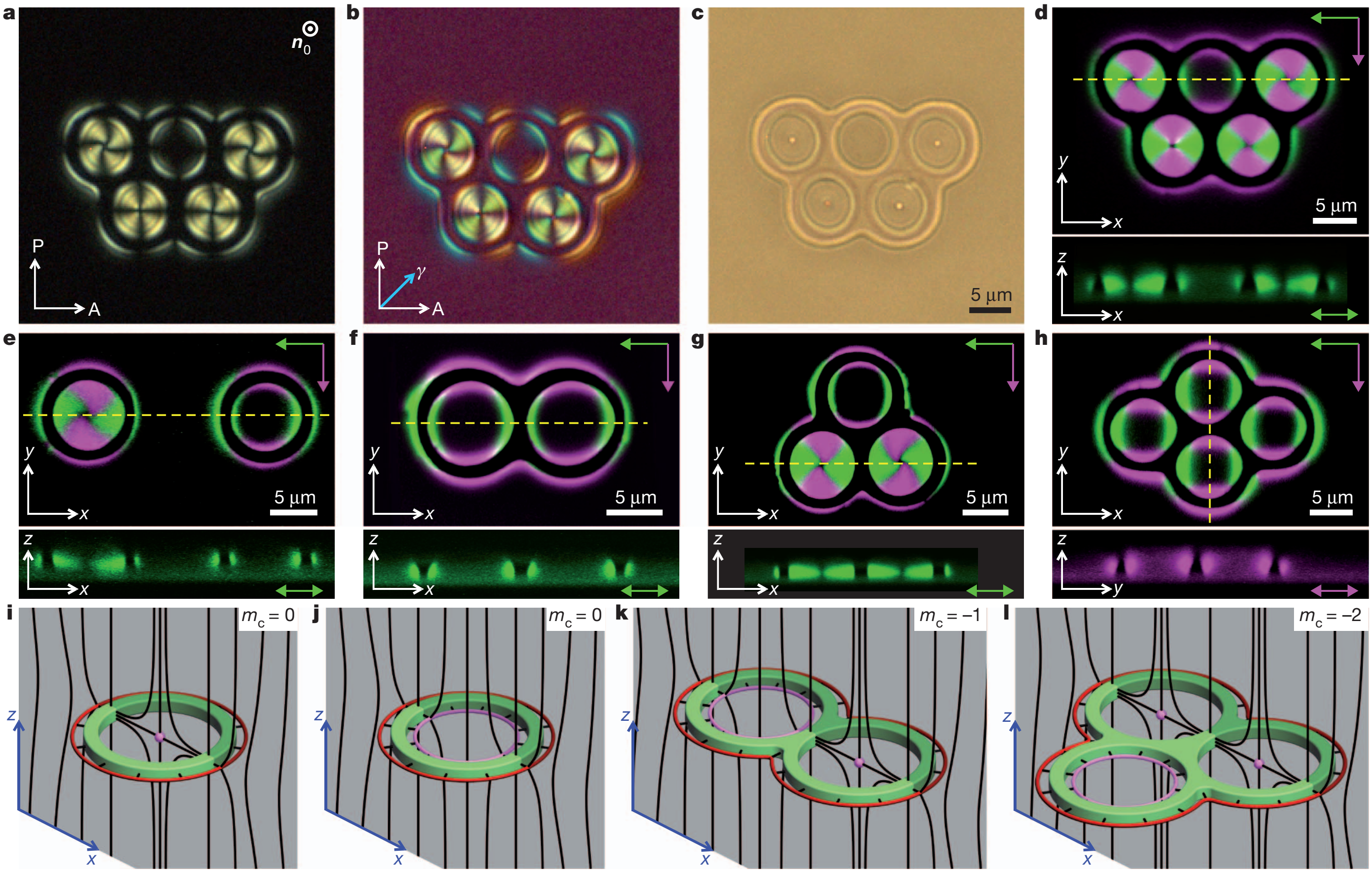


Figure 1 | Colloidal handlebodies aligned orthogonal to the far-field director. **a–d**, $g = 5$ handlebodies and induced $\boldsymbol{n}(\boldsymbol{r})$ structures imaged by PM without (**a**) and with (**b**) a retardation plate, bright-field microscopy (**c**) and 3PEF-PM (**d**) techniques. **e–h**, 3PEF-PM textures of single (**e**), double (**f**), triple (**g**) and quadruple (**h**) handlebodies. P, A and γ mark the crossed polarizer, analyser and a slow axis of a retardation plate (aligned at 45° to P and A), respectively. In the images (**d–h**) obtained by overlaying 3PEF-PM fluorescence intensity patterns for two orthogonal polarizations of excitation light, green and magenta colours correspond to the polarization directions marked by green and magenta arrows, respectively. Cross-sectional *xz* and *yz* images were obtained along yellow lines shown on the corresponding in-plane images. **i–l**, Diagrams of $\boldsymbol{n}(\boldsymbol{r})$ (black lines) around *g* handlebodies. Red and magenta lines show outer and inner disclination loops of $m = -1$ and $m = +1$ hedgehog charges, respectively. Magenta spheres show $m = +1$ hyperbolic point defects.

into point defects and vice versa by melting the liquid crystal into an isotropic state with tweezers of laser power more than 100 mW and subsequently quenching into a nematic phase, indicating that free energies due to director configurations with these defects are comparable. These hedgehog charges of the point defects and disclination loops have been determined by assuming that the vector field lines point perpendicularly outwards from the particle surfaces (Fig. 3) and by mapping the vector fields around particles, point defects and disclination loops onto the order-parameter space[31,32]. Because $\boldsymbol{n}$ has nonpolar symmetry (that is, $\boldsymbol{n}$ is equivalent to $-\boldsymbol{n}$), one could have chosen the vector field pointing inwards to the surface of colloids, which would consequently reverse the signs of all hedgehog charges induced by particles in a uniformly aligned liquid crystal. The relative charges of all the defects would remain the same, as would the net charge of 0, ensuring conservation of topological charge.

Colloidal particles oriented with their rings parallel to $\boldsymbol{n}_0$ tend to induce point defects both within the holes and next to the particles (Fig. 2). The point defects occasionally open into disclination loops that follow the curved edge faces of particles and have a topological charge equivalent to that of the point defect that they replace (Fig. 2b, e, h, k). Although the handlebodies oriented perpendicular and parallel to $\boldsymbol{n}_0$ induce a different director field $\boldsymbol{n}(\boldsymbol{r})$, the sum of hedgehog charges due to induced point defects and disclination loops, $\sum_i m_i = -m_c = \pm\chi/2$, compensates for the colloidal particles' hedgehog charge m_c due to $\boldsymbol{n}(\boldsymbol{r})$ at their surfaces and is uniquely predetermined by particle topology (Fig. 3). The signs depend solely on the choice of the direction of the vector field at the surface of particles. This relation holds for all colloidal handlebodies ($g = 1, 2, \ldots, 5$) and for spherical colloids with $g = 0$ and $\sum_i m_i = \pm\chi/2 = \pm 1$ (Supplementary Fig. 12) studied previously[15], and can be understood using simple considerations based on the Gauss–Bonnet theorem[12]. Recall that the topological charge m_c of any region of space V bounded by a surface $S = \partial V$ is the degree[11–13] of $\boldsymbol{n}$ along S, which can be calculated by integrating the Jacobian of $\boldsymbol{n}(\boldsymbol{r})$ over that surface[10,32,33], $m_c = (1/4\pi)\int_S \mathrm{d}x_1\mathrm{d}x_2\,\boldsymbol{n}\cdot\partial_1\boldsymbol{n}\times\partial_2\boldsymbol{n}$. Because $\boldsymbol{n}(\boldsymbol{r})$ aligns with the (outer) unit normal field to the colloidal surface S, the integral reduces to the total Gauss curvature of S divided by 4π. The Gauss–Bonnet theorem[12] states that the total Gauss curvature of a closed surface without boundary is quantized in units of 4π equal to $4\pi(1-g) = 2\pi\chi$ and remains unchanged during all continuous deformations of the surface; it follows that the hedgehog charge m_c of $\boldsymbol{n}(\boldsymbol{r})$ along S is (up to sign) $m_c = \pm 2\pi\chi/(4\pi) = \pm(1-g)$. Because the director is roughly constant (along $\boldsymbol{n}_0$) far from the colloidal inclusions, an imaginary surface surrounding the colloids and all other defects will have a net zero charge. It follows that the sum of the defect charges must cancel the degree on the colloidal surface S, and so the total hedgehog charge of point defects and disclination loops will be $\sum_i m_i = -m_c = \pm\chi/2 = \pm(1-g)$, regardless of the orientation of the particles with respect to $\boldsymbol{n}_0$, as observed experimentally.

The diagram in Fig. 3g shows that both interior and exterior disclination loops of the configurations shown in Fig. 1 can be transformed to hyperbolic point defects of equivalent hedgehog charge.

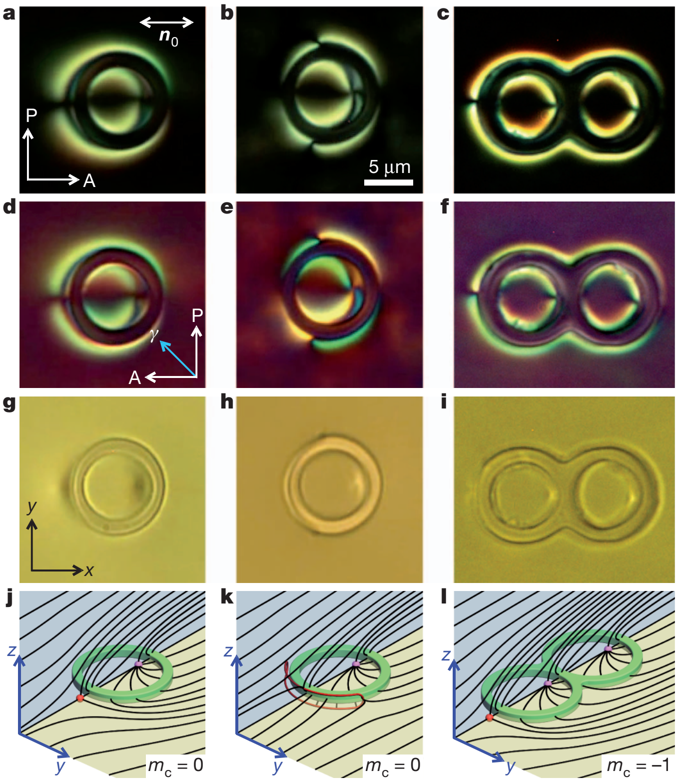


Figure 2 | Colloidal *g* handlebodies aligned parallel to the far-field director. **a–i**, Polarizing (**a–f**) and bright-field (**g–i**) textures and corresponding diagrams of $\boldsymbol{n}(\boldsymbol{r})$ for different colloidal tori. Magenta and red spheres show the $m = +1$ and $m = -1$ hyperbolic point defects, respectively. **j–l**, The red loop in **k** shows a curved half-integer disclination ring with hedgehog charge $m = -1$ observed when a hyperbolic point defect near a solid torus (**a**, **j**) opens into a disclination loop (**b**, **k**). The black lines in **j–l** depict $\boldsymbol{n}(\boldsymbol{r})$ in the plane of colloidal handlebodies (yellow) and in the plane orthogonal to the handlebodies (blue), both planes being parallel to $\boldsymbol{n}_0$.

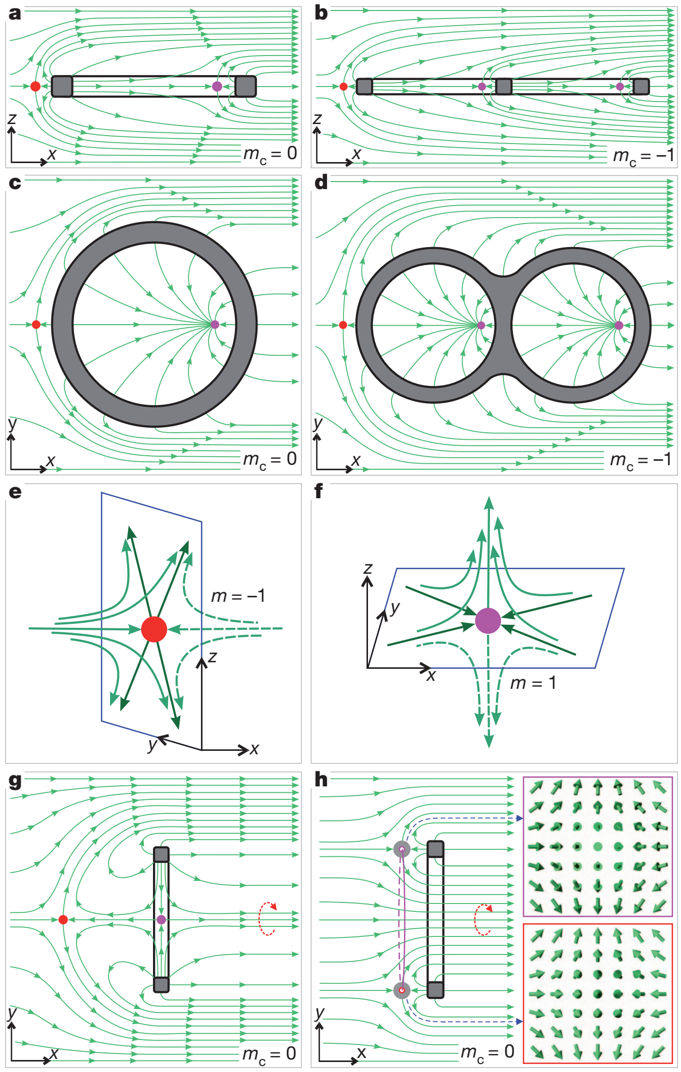


Figure 3 | Director field and point defects around colloidal handlebodies. **a–d**, Diagrams showing the vector-field representations of $\boldsymbol{n}(\boldsymbol{r})$ (green lines with arrows) around single (**a**, **c**) and double (**b**, **d**) colloidal handlebodies in the plane of the rings (**c**, **d**) and in planes orthogonal to them (**a**, **b**). **e**, **f**, Diagrams of $\boldsymbol{n}(\boldsymbol{r})$ around hyperbolic point defects of negative (**e**) and positive (**f**) topological signs shown by red and magenta filled circles. **g**, **h**, Diagrams showing elastic-energy-costly unstable $\boldsymbol{n}(\boldsymbol{r})$ structures with point defects of opposite $m = \pm 1$ (**g**) hedgehog charges and non-singular twist-escaped configuration (**h**) near a handlebody oriented perpendicular to $\boldsymbol{n}_0$. The insets in **h** show the detailed vector field of the escaped axially symmetric configuration corresponding to non-singular $\boldsymbol{n}(\boldsymbol{r})$ of $m = 0$. The cross-section of the field around a handlebody resembles that of an integer-strength disclination loop and is compensated for by an integer-strength disclination loop of opposite strength; the singularity of the latter is removed by an 'escape in the third dimension' by means of continuous deformations as shown in the insets to **h**.

Furthermore, these structures can be also transformed into a non-singular twist-escaped looped $\boldsymbol{n}(\boldsymbol{r})$ configuration with a net topological hedgehog charge equal to zero (Fig. 3h) and resembling the 'bubble gum' structure studied previously[21,22]. Although perpendicular boundary conditions due to the handlebody-shaped particles in the liquid crystal with a uniform $\boldsymbol{n}_0$ can be satisfied by a minimum number of point or ring defects of the same sign having the total hedgehog charge of $\pm\chi/2$ (that is, no singularities for a solid torus, as shown in Fig. 3h, and $g - 1$ point or ring defects for a handlebody of genus g), these field configurations relax to topology-satisfying $\boldsymbol{n}(\boldsymbol{r})$ that also minimize the free energy. The energetic cost of introducing colloids into liquid crystal is dominated by the elastic energy $F = \frac{1}{2}\int \{K_1(\nabla\cdot\boldsymbol{n})^2 + K_2(\boldsymbol{n}\cdot\nabla\times\boldsymbol{n})^2 + K_3(\boldsymbol{n}\times\nabla\times\boldsymbol{n})^2\}\mathrm{d}^3\boldsymbol{r}$, where K_1, K_2 and K_3 are splay, twist and bend elastic constants, respectively, although the total energy additionally includes the surface energy due to finite surface anchoring of $\boldsymbol{n}(\boldsymbol{r})$ at the particle surfaces, the contribution of flexoelectric terms, and the energy of defect cores that can be treated as having a reduced order parameter or a biaxial nature[33]. The surfaces of handlebody colloids have regions with opposite curvature, thus inducing the corresponding distortions of $\boldsymbol{n}(\boldsymbol{r})$ that minimize elastic energy for perpendicular boundary conditions at their surface. This results in the appearance of additional self-compensating pairs of defects of opposite hedgehog charge, leading to a number of defects that exceeds the minimum number, $g - 1$, required by topology. In the experimentally studied systems, colloidal g-handlebodies typically induce $g + 1$ individual singularities. Of these, $g - 1$ defects are of the same charge and are dictated by the particle topology, and two additional defects with opposite signs (total hedgehog charge zero) appear to relax $\boldsymbol{n}(\boldsymbol{r})$ distortions to minimize the free energy. Because $K_2 < K_1 < K_3$ for pentyl cyanobiphenyl[18], some of the splay distortions in the holes of the handlebodies confined into thin cells transform into more complex configurations, as demonstrated by spiralling dark and bright brushes in PM and 3PEF-PM images (Fig. 1a–d, g and Supplementary Figs 3 and 8). Although the structures shown in Fig. 3g, h and Supplementary Fig. 11 are unstable because of their high free energy and are found to relax to other topologically equivalent stable and metastable configurations around individual colloidal handlebodies (Figs 1 and 2), there is a possibility that they could be stabilized by confinement in twisted liquid crystal cells, as

previously observed for the 'bubble gum' configurations formed around colloidal dimers[22].

We have characterized the Brownian motion of colloidal handlebodies (Fig. 4a–c). Their diffusion in a planar cell with thickness much larger than the diameter of the handlebody is highly anisotropic (Fig. 4a, d) and easier along $\boldsymbol{n}_0$ than perpendicular to it. The slopes of mean square particle displacements (MSDs), shown in Fig. 4a for a solid torus ($g = 1$), yield diffusion coefficients $D_x = 0.0023\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$ and $D_y = 0.0034\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$ measured normal and parallel to $\boldsymbol{n}_0$, respectively. Being oriented with respect to $\boldsymbol{n}_0$ (Fig. 4d), particles also experience angular thermal fluctuations (Fig. 4b) with $\langle\Delta\theta^2\rangle$ of angular displacements (MSD_θ) initially increasing linearly with the lag time τ and then saturating as a result of the elasticity-mediated alignment. The width of the histogram distribution of the angle θ between the axis of revolution of the solid torus and $\boldsymbol{n}_0$ (Fig. 4b, inset) is 9.6×10^{-3} rad. The lateral diffusion of these colloids along directions perpendicular to $\boldsymbol{n}_0$ in homeotropic nematic cells is isotropic (Fig. 4c). However, when characterized in the particle's body frame, diffusion of $g > 1$ handlebodies is anisotropic because of their shape. For example, $g = 2$ particles diffuse more easily along their long axis a (Fig. 4c) crossing the centres of the two holes than along the short axis $b \perp a$ (ref. 27). The average diffusion of g-handlebodies having the same diameter of rings decreases with increasing g (Fig. 4c). While being elastically trapped in the vicinity of the handlebodies, defects accompanying the particles also undergo thermal fluctuations.

In addition to laser tweezing and local melting, the relation between defects in $\boldsymbol{n}(\boldsymbol{r})$ and χ can also be probed by applying an electric field $\boldsymbol{E}$ (Fig. 4d) that causes the rotation of $\boldsymbol{n}$ towards $\boldsymbol{E}$ as a result of the liquid crystal's positive dielectric anisotropy. Two types of response have been observed. When $\boldsymbol{E}$ is increased continuously, colloidal handlebodies reorient while preserving their alignment with respect to the director and following its reorientation towards $\boldsymbol{E}$ normal to the

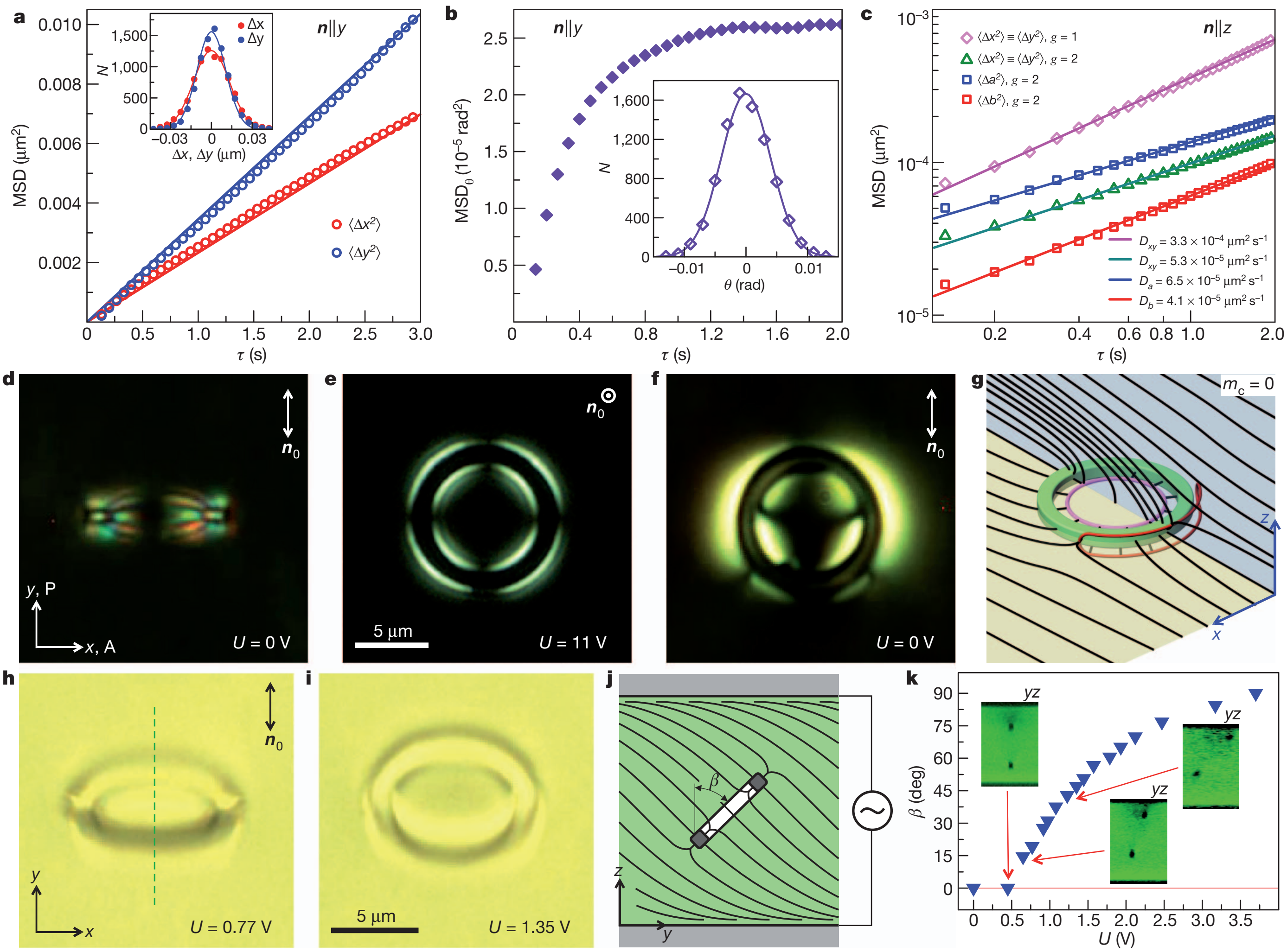


Figure 4 | Diffusion and electric control of colloidal handlebodies in a nematic liquid crystal. **a**, MSD of a solid torus parallel (blue circles) and perpendicular (red circles) to $\boldsymbol{n}_0$ versus τ in a planar nematic cell of thickness $d = 16\,\mu$m. The inset shows the corresponding experimental displacement histograms (red and blue filled circles) at $\tau = 0.067$ s, and Gaussian fits (red and blue lines). **b**, Plot of MSD_θ of a colloidal torus against τ and a histogram of angular orientations with respect to $\boldsymbol{n}_0$ (inset). **c**, MSDs of $g = 1$ and $g = 2$ handlebodies in a homeotropic nematic cell of $d = 3\,\mu$m along x, y, a and b axes; solid lines are linear fits to the data. Note the decreased mobility of handlebodies due to the thin cell confinement. **d–f**, PM textures of a planar nematic cell of $d = 17.5\,\mu$m in the vicinity of a colloidal solid torus at no field (**d**), after continuous increase of applied alternating-current voltage to $U = 11$ V (1 kHz) within about 10 s (**e**), and after voltage $U = 14.8$ V was applied to the cell abruptly (as a square pulse modulated at 1 kHz) and then switched off (**f**). The image in **f** was taken at no applied field and after the torus and $\boldsymbol{n}(\boldsymbol{r})$ had relaxed to the long-lived metastable state with the torus parallel to $\boldsymbol{n}_0$. **g**, Diagram of $\boldsymbol{n}(\boldsymbol{r})$ (black lines) in a texture shown in **f**. **h**, **i**, Bright-field images of a torus in a planar nematic cell of $d = 17.5\,\mu$m reorienting with the liquid crystal director under the applied field $\boldsymbol{E}$ normal to the image. **j**, Diagram showing U-controlled $\boldsymbol{n}(\boldsymbol{r})$ deformations and rotation of the torus in the vertical yz plane. **k**, Plot of torus tilt angle β against U; the insets show 3PEF-PM cross-sectional images along the green dashed line marked in **h** at corresponding U.

substrates (Fig. 4e). However, because of slow rotation of the ring compared with the roughly 10-ms response time of $\boldsymbol{n}(\boldsymbol{r})$, an abrupt application of $\boldsymbol{E}$ simply alters $\boldsymbol{n}(\boldsymbol{r})$ around the particle while preserving the initial particle alignment in the cell. For a solid torus ($g=1$), this causes the original $\boldsymbol{n}(\boldsymbol{r})$ to transform into a topologically equivalent configuration with two disclination loops (Fig. 4f, g). Using different voltage-driving schemes, colloidal handlebodies and structures around them can be switched between the two bistable orientations and $\boldsymbol{n}(\boldsymbol{r})$ configurations shown in Fig 4d, f that are stable at no applied field. All observed transformations of $\boldsymbol{n}(\boldsymbol{r})$ and orientations of colloids are again found to satisfy the relation $\sum_i m_i = \pm\chi/2$.

Our study experimentally supports the procedure for assignment of signs of topological defects in three-dimensional $\boldsymbol{n}(\boldsymbol{r})$ textures before their summation that requires a global point of reference, a 'base point', that serves as a global choice for the overall sign of $\boldsymbol{n}(\boldsymbol{r})$, held fixed during any smooth deformation of the director complexion. Once fixed, the non-polar director field can be decorated with a vector, and the use of vector-field lines allows us to assign unambiguous signs to the defects (Fig. 3a–f)[32]. Although used for many decades[10], the convention that all hyperbolic point defects and disclination loops of $-1/2$ strength have charge '-1' whereas all radial defects and disclination loops of $+1/2$ strength have hedgehog charge '$+1$' fails to properly describe the topological charge conservation of hedgehog charges in the studied three-dimensional textures. The base point and the use of vector field lines in the liquid crystal texture until the signs of the hedgehog charges are assigned with respect to this base point allow a proper addition of hedgehog charges to the net charge of $\pm\chi/2$. The signs of topological point defects and the entire $\boldsymbol{n}(\boldsymbol{r})$ structure induced by handlebody colloids then depend on the direction of vector field lines at the base point and can be reversed by flipping this direction to an opposite one, because of the non-polar nature of nematic liquid crystals[33]. It is only through the use of a base point that defects in different places can be added together like charges so that the net topological charge is conserved. Our approach also describes topological charge conservation in liquid-crystal textures studied previously, as we show in the Supplementary Information with an example of a colloidal dimer and surrounding $\boldsymbol{n}(\boldsymbol{r})$. One can assign and add hedgehog charges in $\boldsymbol{n}(\boldsymbol{r})$ of samples with multiple colloidal particles having the same or different χ, and the addition of each separate particle always contributes a net $\pm\chi/2 = -m_c$ to the topological charge distribution of particle-induced bulk defects compensating for m_c and ensuring charge conservation. However, charges induced by particles can have opposite signs even within the same texture, thus enabling charge annihilation in the textures surrounding these particles, as we show in the Supplementary Information with an example of colloidal dimers.

We have designed and fabricated topologically distinct handlebody-shaped colloidal particles and explored the interplay between the topology of colloids and the defects that they induce in a uniformly aligned liquid crystal. These handlebody colloids are accompanied by topological defects with the net hedgehog charge always equal to half of the Euler characteristic of the particle surface. Topological colloids and the established procedure for the assignment and summation of topological charges in liquid crystals will enable basic studies of topological manifolds and the interplay between particle topology and order and disorder[7,33] with these model systems. Beyond the exploration of the topology of colloids, fields and defects, the experimental arena we have developed may enable the design of topology-dictated elastic colloidal interactions and reconfigurable self-assembly in liquid crystals[18], the entrapment and scaffolding of nanoparticles by particle-induced defects[23], the self-assembly of reconfigurable topological memory devices[14], and electro-optic and photonic devices based on bistable switching between different states with distinct director configurations and orientations of particles.

METHODS SUMMARY

Fabrication of silica (SiO_2) particles with handlebody topology involved the following procedure. First, a 90-nm sacrificial layer of aluminium was sputtered on a silicon wafer. Next, a 1-μm silica layer was deposited on the aluminium by plasma-enhanced chemical vapour deposition. Photoresist AZ5214 (Clariant AG) was spin-coated on the silica layer. The pattern of rings was defined in the photoresist by illumination at 405 nm with a direct laser-writing system (DWL 66FS; Heidelberg Instruments) and then in the silica layer by inductively coupled plasma etching. Finally, the photoresist was removed with acetone and the aluminium was wet-etched with sodium hydroxide aqueous solution so that the handlebody particles were released and then re-dispersed in deionized water (Supplementary Fig. 1). To define perpendicular boundary conditions for $\boldsymbol{n}(\boldsymbol{r})$ on the surface of particles, they were treated with an aqueous solution (0.05wt%) of *N*,*N*-dimethyl-*N*-octadecyl-3-aminopropyl-trimethoxysilyl chloride (DMOAP) and then re-dispersed in methanol. After the addition of pentyl cyanobiphenyl and the evaporation of methanol at 70 °C overnight, the ensuing nematic dispersion was infiltrated into cells composed of indium–tin-oxide (ITO)-coated glass plates separated by glass spacers defining the cell gap. Cell substrates were treated with DMOAP to achieve perpendicular $\boldsymbol{n}_0$ or coated with polyimide PI2555 (HD Microsystems) for in-plane alignment of $\boldsymbol{n}_0$ defined by rubbing. Optical manipulation and three-dimensional imaging of samples were performed with an integrated setup of holographic optical tweezers and 3PEF-PM (Supplementary Fig. 2)[23,30] built around an inverted microscope IX 81 (Olympus) and using a 100× oil-immersion objective (numerical aperture 1.4). Holographic optical tweezers used a phase-only spatial light modulator (Boulder Nonlinear Systems) and an ytterbium-doped fibre laser (IPG Photonics) operating at 1,064 nm. 3PEF-PM employed a tunable (680–1,080 nm) Ti–sapphire oscillator (Coherent) emitting 140-fs pulses at a repetition rate of 80 MHz, and a photomultiplier tube detector H5784-20 (Hamamatsu)[30].

Received 2 September; accepted 24 October 2012.
Published online 23 December 2012.

1. Hunter, R. J. *Foundations of Colloid Science* (Oxford Univ. Press, 2001).
2. Alsayed, A. M., Islam, M. F., Zhang, J., Collings, P. J. & Yodh, A. G. Premelting at defects within bulk colloidal crystals. *Science* **309,** 1207–1210 (2005).
3. Semonin, O. E. *et al.* Peak external photocurrent quantum efficiency exceeding 100% via MEG in a quantum dot solar cell. *Science* **334,** 1530–1533 (2011).
4. Hud, N. V., Downing, K. H. & Balhord, R. A constant radius of curvature model for the organization of DNA in toroidal condensates. *Proc. Natl Acad. Sci. USA* **92,** 3581–3585 (1995).
5. Michalet, X. & Bensimon, D. Observation of stable shapes and conformal diffusion in genus 2 vesicles. *Science* **269,** 666–668 (1995).
6. Hsu, L., Kusner, R. & Sullivan, J. Minimizing the squared mean curvature integral for surfaces in space forms. *Exp. Math.* **1,** 191–207 (1992).
7. Kamien, R. D. The geometry of soft materials: a primer. *Rev. Mod. Phys.* **74,** 953–971 (2002).
8. Lavrentovich, O. D. & Terentjev, E. M. Phase transition altering the symmetry of topological point defects (hedgehogs) in a nematic liquid crystal. *Sov. Phys. JETP* **64,** 1237–1244 (1986).
9. Tkalec̆, U., Ravnik, M., Čopar, S., Žumer, S. & Muševič, I. Reconfigurable knots and links in chiral nematic colloids. *Science* **33,** 62–65 (2011).
10. Kleman, M. & Lavrentovich, O. D. Topological point defects in nematic liquid crystals. *Phil. Mag.* **86,** 4117–4137 (2006).
11. Montiel, S. & Ros, A. *Curves and Surfaces* 2nd edn (American Mathematical Society, 2009).
12. Milnor, J. W. *Topology from the Differentiable Viewpoint* (Univ. Press of Virginia, 1965).
13. Hopf, H. Vektorfelder in *n*-dimensionalen Mannigfaltigkeiten. *Math. Ann.* **96,** 225–249 (1927).
14. Araki, T., Buscaglia, M., Bellini, T. & Tanaka, H. Memory and topological frustration in nematic liquid crystals confined in porous materials. *Nature Mater.* **10,** 303–309 (2011).
15. Poulin, P., Holger, S., Lubensky, T. C. & Weitz, D. A. Novel colloidal interactions in anisotropic fluids. *Science* **275,** 1770–1773 (1997).
16. Irvine, W. T. M., Vitelli, V. & Chaikin, P. M. Pleats in crystal on curved surfaces. *Nature* **468,** 947–951 (2010).
17. Jones, M. R. *et al.* DNA-nanoparticle superlattices formed from anisotropic building blocks. *Nature Mater.* **9,** 913–917 (2010).
18. Lapointe, C. P., Mason, T. G. & Smalyukh, I. I. Shape-controlled colloidal interactions in nematic liquid crystals. *Science* **326,** 1083–1086 (2009).
19. Sacanna, S., Irvine, W. T. M., Chaikin, P. M. & Pine, D. J. Lock and key colloids. *Nature* **464,** 575–578 (2010).
20. Wood, T. A., Lintuvuori, J. S., Schofield, A. B., Marenduzzo, D. & Poon, W. C. K. A self-quenched defect glass in a colloid–nematic liquid crystal composite. *Science* **334,** 79–83 (2011).

21. Poulin, P., Cabuil, V. & Weitz, D. A. Direct measurement of colloidal forces in an anisotropic solvent. *Phys. Rev. Lett.* **79,** 4862–4865 (1997).
22. Tkalec, U., Ravnik, M., Žumer, S. & Muševič, I. Vortexlike topological defects in nematic colloids: chiral colloidal dimers and 2D crystals. *Phys. Rev. Lett.* **103,** 127801 (2009).
23. Trivedi, R. P., Klevets, I. I., Senyuk, B., Lee, T. & Smalyukh, I. I. Multi-scale interactions and three-dimensional patterning of colloidal particles and defects in lamellar soft media. *Proc. Natl Acad. Sci. USA* **109,** 4744–4749 (2012).
24. Herges, R. Topology in chemistry: designing Möbius molecules. *Chem. Rev.* **106,** 4820–4842 (2006).
25. Mosseri, R. Geometrical frustration and defects in condensed matter systems. *C. R. Chim.* **11,** 192–197 (2008).
26. Bowick, M. J., Chandar, L., Schiff, E. A. & Srivastava, A. M. The cosmological Kibble mechanism in the laboratory: string formation in liquid crystals. *Science* **263,** 943–945 (1994).
27. Han, Y. *et al.* Brownian motion of an ellipsoid. *Science* **314,** 626–630 (2006).
28. Solomon, M. J. Directions for targeted self-assembly of anisotropic colloids from statistical thermodynamics. *Curr. Opin. Colloid Interface Sci.* **16,** 158–167 (2011).
29. Merkel, T. J. *et al.* Scalable, shape-specific, top-down fabrication methods for the synthesis of engineered colloidal particles. *Langmuir* **26,** 13086–13096 (2010).
30. Lee, T., Trivedi, R. P. & Smalyukh, I. I. Multimodal nonlinear optical polarizing microscopy of long-range molecular order in liquid crystals. *Opt. Lett.* **35,** 3447–3449 (2010).
31. Pettey, D. & Lubensky, T. C. Stability of texture and shape of circular domains of Langmuir monolayers. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **59,** 1834–1845 (1999).
32. Alexander, G. P., Chen, B. G., Matsumoto, E. A. & Kamien, R. D. Colloquium: Disclination loops, point defects, and all that in nematic liquid crystals. *Rev. Mod. Phys.* **84,** 497–514 (2012).
33. Chaikin, P. M. & Lubensky, T. C. *Principles of Condensed Matter Physics* (Cambridge Univ. Press, 2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Chen, N. Clark, J.-i. Fukuda and S. Žumer for discussions. This work was supported by the International Institute for Complex Adaptive Matter and the National Science Foundation grants DMR-0844115 (Q.L., S.H. and I.I.S.), DMR-0820579 (B.S. and I.I.S.), DMR-0847782 (B.S., Q.L. and I.I.S.), PHY11-25915 (R.D.K., R.B.K., T.C.L. and I.I.S.) and DMR-1120901 (R.D.K. and T.C.L.). R.B.K., R.D.K., T.C.L. and I.I.S. thank the Kavli Institute for Theoretical Physics for their hospitality while this work was being discussed and prepared for publication.

Author Contributions B.S., Q.L. and I.I.S. performed experimental work. Q.L., S.H. and I.I.S. designed and fabricated particles. B.S. and I.I.S. reconstructed director fields induced by colloids. T.C.L. and I.I.S. characterized topological charges of defects in particle-induced director fields. R.B.K., R.D.K., T.C.L. and I.I.S. proposed models of field transformations satisfying topological constraints and explained the relations between genus of colloids and the net topological charge of liquid crystal defects. I.I.S. conceived the project, designed experiments, provided funding and wrote the manuscript. All authors edited and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.I.S. (ivan.smalyukh@colorado.edu).

LETTER

doi:10.1038/nature11727

Large-scale nanophotonic phased array

Jie Sun[1], Erman Timurdogan[1], Ami Yaacobi[1], Ehsan Shah Hosseini[1] & Michael R. Watts[1]

Electromagnetic phased arrays at radio frequencies are well known and have enabled applications ranging from communications to radar, broadcasting and astronomy[1]. The ability to generate arbitrary radiation patterns with large-scale phased arrays has long been pursued. Although it is extremely expensive and cumbersome to deploy large-scale radiofrequency phased arrays[2], optical phased arrays have a unique advantage in that the much shorter optical wavelength holds promise for large-scale integration[3]. However, the short optical wavelength also imposes stringent requirements on fabrication. As a consequence, although optical phased arrays have been studied with various platforms[4–8] and recently with chip-scale nanophotonics[9–12], all of the demonstrations so far are restricted to one-dimensional or small-scale two-dimensional arrays. Here we report the demonstration of a large-scale two-dimensional nanophotonic phased array (NPA), in which 64 × 64 (4,096) optical nanoantennas are densely integrated on a silicon chip within a footprint of 576 μm × 576 μm with all of the nanoantennas precisely balanced in power and aligned in phase to generate a designed, sophisticated radiation pattern in the far field. We also show that active phase tunability can be realized in the proposed NPA by demonstrating dynamic beam steering and shaping with an 8 × 8 array. This work demonstrates that a robust design, together with state-of-the-art complementary metal-oxide–semiconductor technology, allows large-scale NPAs to be implemented on compact and inexpensive nanophotonic chips. In turn, this enables arbitrary radiation pattern generation using NPAs and therefore extends the functionalities of phased arrays beyond conventional beam focusing and steering, opening up possibilities for large-scale deployment in applications such as communication, laser detection and ranging, three-dimensional holography and biomedical sciences, to name just a few.

An NPA, resembling its radiofrequency counterpart, consists of several typically identical optical antennas. Each antenna emits light of a specific amplitude and phase to form a desired far-field radiation pattern through interference of these emissions. The short wavelength of light offers potential for NPAs to greatly exceed the number of elements found in their radiofrequency counterparts, with thousands or even millions of elements in a compact, low-cost chip. By incorporating a large number of antennas, high-resolution far-field patterns can be achieved and therefore arbitrary radiation patterns can be generated by an NPA, extending the functionalities of phased arrays well beyond the conventional beam focusing and steering. However, the short optical wavelength also presents challenges in realizing coherent outputs from such large-scale NPAs because even nanoscale dimensional fluctuations affect the optical emission from the thousands of nanoantennas that need to be balanced in power and aligned in phase to form a specific far-field radiation pattern. As a consequence, all chip-based two-dimensional NPAs demonstrated so far are limited to small-scale implementations with no more than 16 antennas, and their functionalities are thereby constrained to conventional single-beam focusing and steering.

Here we report the demonstration of a large-scale, compact NPA system that is compatible with a complementary metal-oxide–semiconductor (CMOS) process, consisting of 64 × 64 optical nanoantennas on a silicon chip with all 4,096 optical nanoantennas balanced in power and aligned in phase to produce a sophisticated radiation pattern, the MIT logo, in the far field. To our knowledge, this demonstration represents the largest coherent combination of nanophotonic elements so far. It also shows that despite the short optical wavelength and corresponding length of the phase elements, the phase of the elements can be maintained, highlighting the ability to make arbitrary manipulations of the phase of an optical field within a nanophotonic chip. In addition, we demonstrate that active phase tunability can be successfully implemented in the proposed NPA structure to generate dynamic far-field patterns.

The structure of the 64 × 64 NPA is illustrated schematically in Fig. 1a, and Fig. 1b shows a scanning electron micrograph (SEM) of the fabricated NPA. A laser input is coupled into the main silicon bus waveguide through an optical fibre, and then evanescently coupled into 64 row waveguides. The coupling to the row waveguides is controlled in such a way that each row waveguide obtains the same amount of power by varying the length of the directional coupler (Supplementary Fig. 1). The optical power in each row waveguide is then similarly divided into 64 optical antenna units so that all 4,096 optical nanoantennas are uniformly excited.

Each of the 4,096 optical antenna units, or 'pixels', consists of two major functional parts (Fig. 1a, inset): a nanoantenna to efficiently deliver optical power to free space, and two optical phase delay lines to accurately adjust the phase of the light arriving at the nanoantenna (Supplementary Fig. 1). Figure 1c shows the SEM of the fabricated pixel. Ideally, the pixel size should be less than half of the free-space wavelength λ_0 of the optical emission in both the x and y directions for the radiation to have a unique interference pattern in the far field without high-order radiation lobes[13,14]. It is therefore crucial to make the pixel as small as possible, so as to decrease the number of high-order interference patterns in the far field. A compact and efficient optical nanoantenna is thus demanding for the NPA system. Optical nanoantennas down to subwavelength size have previously been demonstrated[15–19] with metal; however, to make the material systems compatible with the CMOS process, silicon-based dielectric nanoantennas are used (Fig. 1c). The compact grating-based optical nanoantenna measures 3.0 μm in length, 2.8 μm in width, and 0.22 μm in thickness, consisting of only five grating teeth. A partial etch 0.11 μm in depth is applied to the first grating groove, to break the up–down symmetry of the nanoantenna for more upward emission[20]. Figure 2a shows the simulated near-field emission pattern from a three-dimensional finite-difference time-domain simulation, in which 51% of the optical power emits upwards while only 30% emits downwards at $\lambda_0 = 1.55$ μm. Figure 2b illustrates the calculated far-field pattern of the optical nanoantenna, using the near-to-far-field transformation[21]. The emission is not vertical because the grating period is slightly detuned from that of a second-order grating that would emit vertically, to suppress the resonant back-reflections; otherwise, significant reflection would interfere with the light propagation in the phased array. The emission from the nanoantenna is also broadband, with a bandwidth extending across hundreds of nanometres in wavelength, an inherent

[1]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

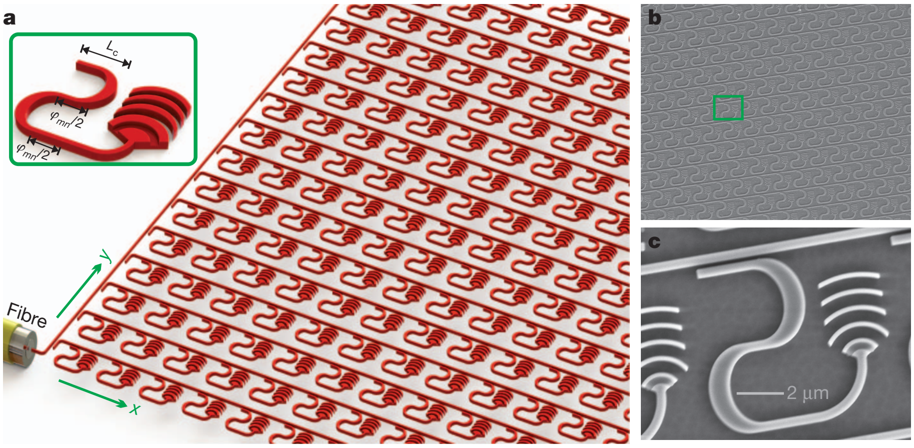


Figure 1 | The NPA system. **a**, Schematic illustration of a 64 × 64 NPA system. Laser input from an optical fibre is delivered equally to each of the 4,096 nanoantennas through silicon waveguides. The inset shows a diagram of a close-up view of one antenna unit cell (m, n), or 'pixel'. The coupling efficiency is varied by the length of the directional coupler L_c, and the emitting phase is controlled by two segments of optical delay lines to achieve a desired phase delay of φ_{mn}. **b**, SEM of part of the 64 × 64 NPA system fabricated at a CMOS foundry. **c**, A close-up SEM of one pixel of the NPA system, indicated by the green rectangle in **b**. The pixel size is 9 μm × 9 μm, with a compact grating as an optical nanoantenna, where the first groove of the grating is partly etched to enhance the upward emission. The emitted phase of each pixel is adjusted by the length of the optical delay line within the pixel.

characteristic of the short length of the grating (see also Supplementary Fig. 2).

As mentioned above, the ability to integrate a large number of pixels in the NPA within a small footprint opens up the possibility of using NPA to generate arbitrary, sophisticated far-field radiation patterns. The far-field radiation field $E(\theta,\phi)$ of the phased array is calculated as the far field of an individual nanoantenna $S(\theta,\phi)$ (Fig. 2b), multiplied by the array factor $F_a(\theta,\phi)$[22]

$$E(\theta,\phi) = S(\theta,\phi) \times F_a(\theta,\phi) \tag{1}$$

where the array factor is a system factor that is related to the phase of optical emission from all the pixels. θ and ϕ are the far-field azimuth angle and polar angle, respectively. By assigning the optical phase φ_{mn} of each pixel (where m and n are the pixel indices) in the NPA, the desired radiation pattern $E(\theta,\phi)$ can be achieved. The phase φ_{mn} of each pixel can be determined by antenna synthesis through the Gerchberg–Saxton algorithm[23,24] (Supplementary Fig. 3). The background pattern in Fig. 2c shows the simulated array factor pattern of the 64 × 64 NPA at a wavelength of 1.55 μm, designed to generate the MIT logo in the far field. The pixel pitch is chosen to be 9 μm in both the x and y directions, as will be used in fabrication. The pixel pitch is a multiple of the free-space half-wavelength. The designed pattern is therefore replicated in the far field, as shown by the background pattern in Fig. 2c, because the same interference conditions occur periodically in the far field to produce higher-order patterns. Equation (1) indicates that the final far-field radiation pattern of the NPA is a multiplication of the far field of the individual nanoantenna $S(\theta,\phi)$ (Fig. 2b) and that of the array factor $F_a(\theta,\phi)$ (background pattern in Fig. 2c), as shown in Fig. 2c. It is seen that the designed pattern is only visible in the vicinity of the zenith, owing to the directional emission of the optical nanoantenna. Figure 2d shows a close-up view of the far field radiation pattern around the zenith, where only the top rows have enough intensity to be observed.

The simulation shows that, in principle, arbitrary radiation patterns can be produced in the far field with large-scale NPAs, by controlling the emitted phase of all the pixels correctly. However, observing the short optical wavelength (1.55 μm) and the high refractive index of silicon ($n = 3.48$), it is almost impossible to provide every pixel exactly with the designed phase φ_{mn}, because slight fabrication imperfections will cause significant phase errors. As a consequence, only if the designed NPA system is highly resistant to phase errors can this large-scale NPA be fabricated reliably and function properly.

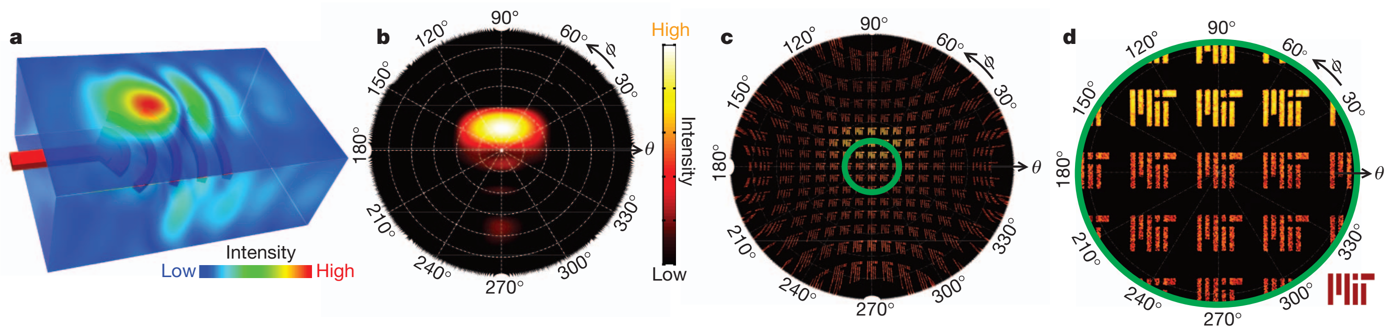


Figure 2 | Device and system simulations. **a**, Three-dimensional near-field emission simulation of the designed optical nanoantenna using the finite-difference time-domain method. **b**, The corresponding far-field radiation pattern of the optical nanoantenna, calculated from the near-field emission (in **a**), using the near-to-far-field transformation. **c**, The radiation pattern of the designed 64 × 64 NPA system to generate the MIT logo in the far field, which is a superposition of the far field of the system's array factor (as shown in the background) and that of the nanoantenna (in **b**). The green circle in the centre indicates the viewable region in a microscope lens (numerical aperture = 0.4) used later in the experiment (Fig. 3a). **d**, A close-up view of the viewable region of the far field displaying the MIT logo. The inset on the lower right shows the targeted MIT logo pattern. The radiation fields in **b**–**d** are viewed from the zenith of the far-field hemisphere, as a projection of the far-field hemisphere to the equatorial plane in the polar coordinate system (θ,ϕ). θ and ϕ are the far-field azimuth angle and polar angle, respectively.

Fortunately, the designed large-scale NPA system is indeed highly tolerant of phase errors (Supplementary Fig. 4). This high phase-error tolerance of the designed NPA system originates from its nature as a Fourier system, in which the phase noise of the near-field emission averages out in the far field through interference of optical emissions from all the pixels. This high phase-error tolerance therefore becomes more effective with more pixels and will thus enable NPAs to scale up beyond 64 × 64, even to millions of pixels.

The robustness of the designed NPA demonstrates such a system can be fabricated reliably. The designed 64 × 64 NPA was fabricated in a 300-mm CMOS foundry by 193-nm optical immersion lithography (Methods). The 4,096 pixels were accurately reproduced, as shown by the SEMs in Fig. 1b, c. A 1.55-μm laser was coupled into the transverse electric mode of the main bus waveguide. Then the near field and far field images were captured with an infrared camera, using a setup (Fig. 3a) similar to that in ref. 25. Figure 3b shows the near-field image of the fabricated NPA system when light was coupled in, where uniform optical emission was observed across all 4,096 nanoantennas. Figure 3c shows a close-up view containing 8 × 8 pixels. It is crucial for an NPA to have all the pixels emit with a desired amplitude pattern, which is uniform here, to provide an ideal interference condition in the far field. Figure 3d quantitatively analyses the uniformity of the optical emission by measuring the near-field emission intensity of each pixel with the infrared camera. A high degree of uniformity is observed. The top-left corner of the NPA is noisier, because it is close to the fibre coupling point where strong scattering occurs, but this does not reflect non-uniformity in the array itself and can readily be addressed with a larger separation from the fibre input.

Figure 3e shows the measured far-field radiation pattern of the fabricated 64 × 64 NPA. The image reveals that the designed sophisticated radiation pattern (MIT logo) is generated in the far field. The far-field image is clamped by the finite numerical aperture (0.4) of lens 1. This is also predicted by simulations, as shown by the green circle in Fig. 2c, d, where only emission with a small divergence angle from vertical can be captured. The intensity noise in the background of the far-field image comes from the light scattering caused by fibre-to-waveguide coupling. The scattered light is also responsible for the concentric fringes in the background, through the interference of the scattered light between the top and bottom surfaces of the silicon-on-insulator wafer. This noise can be removed by placing the fibre–waveguide coupler farther from the NPA system to reduce the light scattering captured by the imaging column, and a much cleaner far-field radiation pattern would be expected. Figure 3f shows the captured radiation pattern of a 32 × 32 NPA on the same chip, in which less noise is observed because it is placed a little farther from the fibre coupling point; however, the far-field pattern resolution is reduced because it contains fewer pixels. The measured images agree perfectly with simulations in Fig. 2c, d, in terms of the shape of the pattern (MIT logo) and the relative intensity of all interference orders, highlighting the robustness of the NPA design and the accuracy of the fabrication. Comparing Fig. 3b with Fig. 3e, the near-field image of the NPA contains no information because it shows plain uniform

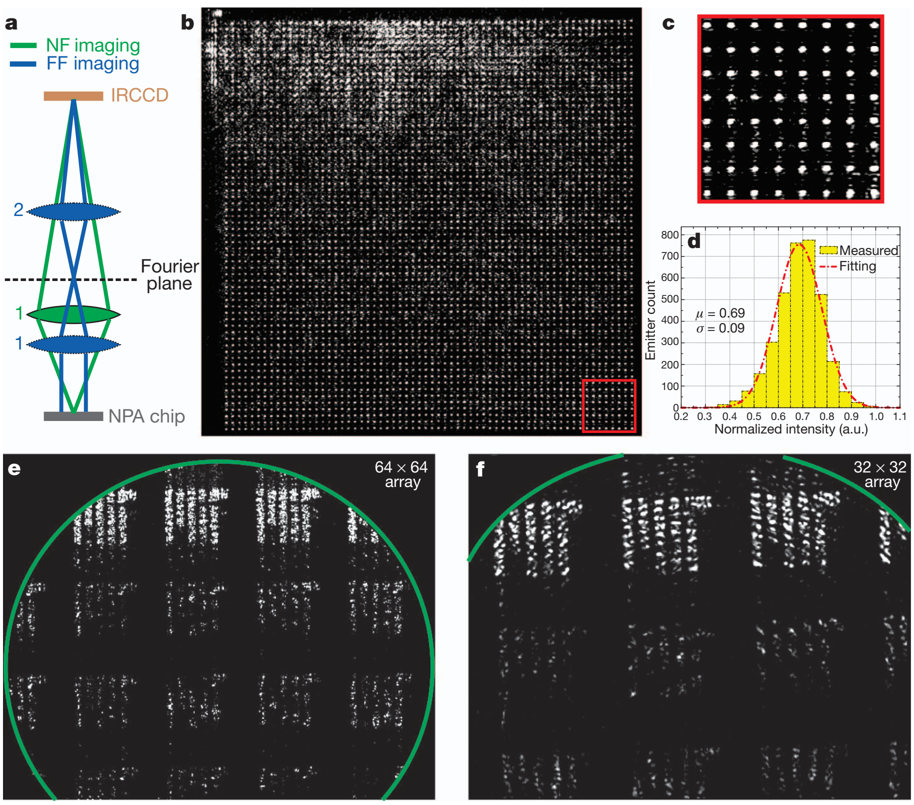


Figure 3 | Experimental results. **a**, A diagram of the imaging system used to observe the near field and far field. Lens 1 alone (numerical aperture = 0.40) was used to obtain the near-field (NF) image, as shown by the green rays. The far-field (FF) image, or the Fourier image, was taken by moving lens 1 down so as to form the far-field image in its back-focal plane (Fourier plane), and inserting lens 2 to project the far-field image on the infrared charge-coupled device (CCD), as shown by the blue rays. **b**, The near-field image shows uniform emission across all of the 64 × 64 (4,096) nanoantennas. The input bus waveguide is located on the top left corner, causing some excess scattering noise. **c**, A close-up view of part of the near field, containing 8 × 8 pixels. **d**, Measured intensity distribution of the optical emission from the pixels. The statistics show that the standard deviation (σ) of the emission intensity is 13% of the average intensity (μ). **e**, The far-field radiation pattern of the fabricated 64 × 64 NPA system. The aperture (indicated by the green circle) corresponds to the numerical aperture (0.4) of lens 1 in **a**. The measurement accurately matched simulations in Fig. 2c, d, except for the higher background noise introduced by the light scattered from fibre-to-waveguide input coupling. **f**, The far-field radiation pattern of a 32 × 32 NPA on the same chip. Less noise but lower resolution is observed. The concentric ring pattern superimposed in the image is caused by the interference of the scattered light between the top and bottom surfaces of the silicon-on-insulator wafer.

emission everywhere; however, the far field explicitly delivers the information—an image with the MIT logo. Until now, image information has generally been stored and transmitted through the intensity of the pixels; in contrast, this large-scale NPA technology opens up another dimension for imaging: the image information is now encoded in the optical phase of the pixels, much like a hologram, but generated from a single point. This demonstration, as a static phased array capable of generating truly arbitrary radiation patterns, would also find immediate applications in, for example, complex beam generation[26] and mode matching in optical space-division multiplexing[27].

By comparison with other holographic approaches such as the metasurface antennas[12], the NPA allows separate control over the phase and amplitude of light emission and on-chip single-point excitation of the nanophotonic emitters, enabling truly arbitrary holograms to be generated entirely on-chip for the first time; moreover, benefiting from the use of guided light in silicon instead of free-space light, active manipulation of the optical phase can be directly implemented to achieve dynamic far-field patterns with more flexibility and wider applications, by converting the pixel into a thermally phase-tunable pixel in a CMOS process (Fig. 4a, inset). A portion of the silicon light path in each pixel is lightly doped with an n-type implant to form a resistive heater for thermo-optic phase tuning while maintaining a low loss of light propagation. Two narrow silicon leads with heavy n^+-doping, providing electrical connections to and thermal isolation from the heater, are connected to the heater on the inner side of the adiabatic bends to minimize the loss caused by light scattering[28,29]. Figure 4a illustrates an active 8×8 NPA in which each pixel has an independently tunable phase shifter; the electrical controls are connected in rows and in columns to simplify the electrical circuitry. This active 8×8 phased array was fabricated with CMOS-processing techniques (Methods). The measured resistance is 2.5 kΩ per heater including the two copper–silicon contacts, and a high thermal efficiency of about 8.5 mW per π-phase shift is achieved, benefiting from the direct heating of the silicon waveguide. By applying different voltages on each pixel, different phase combinations can be achieved in the phased array to generate different radiation patterns dynamically in the far field, as shown in Fig. 4b–f, in which five different radiation patterns

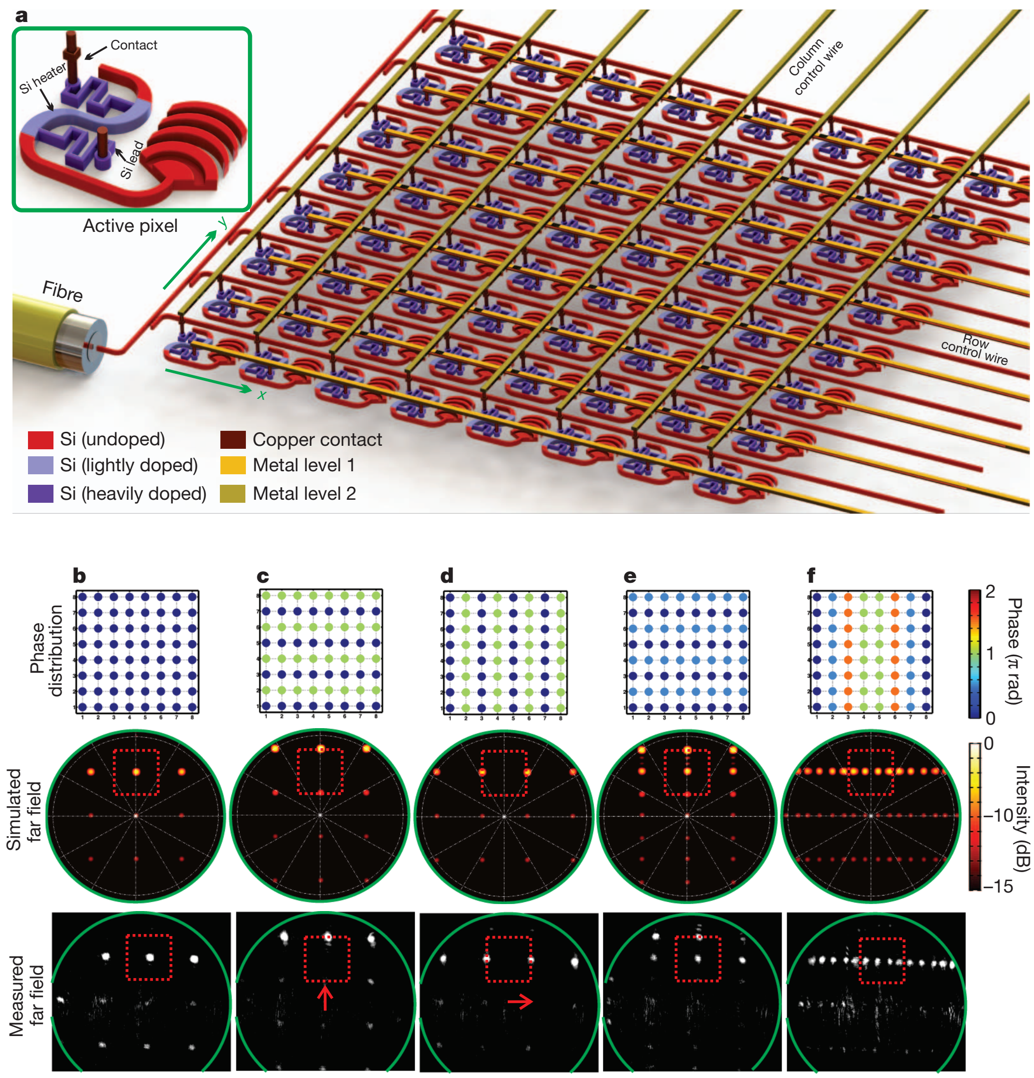


Figure 4 | Tunable phased array. **a**, Schematic illustration of the 8×8 active phased array. Inset, diagram of an active pixel with the same size (9 μm × 9 μm) as the passive pixel in Fig. 1a. The optical phase of each pixel is continuously tuned by the thermo-optic effect through an integrated heater formed by doped silicon. **b**–**f**, Examples of the dynamic far-field patterns generated by the 8×8 active phased array by applying different voltage combinations to the pixels, showing simulations and measurements. **b**, The original single-beam pattern with no voltage on. **c**, **d**, The focused beam is steered by 6° to the edge of each interference order in the vertical (**c**) and in the horizontal (**d**). **e**, The single beam is split into two beams in the vertical direction. **f**, The single beam is split into four beams in the horizontal direction. The green circle indicates the edge of lens 1 (numerical aperture = 0.4), and the red box specifies the area of one interference order (see also the Supplementary Movie for the dynamic pattern generation).

are demonstrated using the same phased array: a standard single-beam pattern with no voltage on the array; beam-steering for about 6° in both the vertical direction and the horizontal direction; generation of two beams in vertical direction; and generation of four beams in the horizontal direction. Good agreement between simulation and experiment is again observed, which confirms the robustness of the proposed design as well as the accuracy of the fabrication and active thermo-optic phase tuning. The proposed active NPA structure can be immediately extended to a larger phased array (for example 64×64, as discussed above) with independent electrical control of each pixel with the aid of fully CMOS-controlled circuitry to access all of the pixels electrically, to project dynamic patterns in the far field with applications ranging from communications, three-dimensional holographic displays, laser detection and ranging (LADAR) and biomedical imaging to interferometry.

We have demonstrated a large-scale 64×64 nanophotonic phased array built on a silicon photonic platform. The robustness and CMOS compatibility of the designed system enabled the successful integration of 4,096 optical nanoantennas within a chip-size footprint, probably the largest and densest silicon photonic functional system demonstrated so far, and certainly the largest coherent demonstration of silicon photonics. Optical power is uniformly distributed in the nanoantennas and is phase aligned to produce a sophisticated far-field radiation pattern, the MIT logo, as designed. The large number of nanoantennas and the embedded phase tunability enable NPAs to generate arbitrary far-field radiation patterns dynamically and, in turn, to affect new fields such as communication, LADAR, three-dimensional holography and biomedical sciences. The ability to take advantage of a state-of-the-art CMOS integration process also promises a bright future for low-cost and compact NPAs that will be important in many fields with their newly explored functionalities.

METHODS SUMMARY

The NPAs were fabricated in a 300-mm CMOS foundry with a 65-nm technology node, using silicon-on-insulator wafers with a 0.22-μm top silicon layer and 2 μm buried oxide. A timed partial silicon etch (0.11 μm) was first performed to make the partly etched grating groove. A full silicon etch was then applied to form the waveguides and grating nanoantennas. Subsequent n and n^+ dopings were implanted for active arrays, followed by standard silicidation to make copper–silicon contacts. The contacts were connected to on-chip probing pads by two metal layers for thermo-optic tuning. SiO_2 with a total thickness of 3.6 μm was used to cover the devices, with a final polishing step to make the surface planar to avoid additional phase errors due to surface corrugation.

Received 2 August; accepted 29 October 2012.

1. Hansen, R. C. *Phased Array Antennas* (Wiley, 1998).
2. Cherry, M. Astronomy in South Africa: the long shot. *Nature* **480,** 308–309 (2011).
3. McManamon, P. F. *et al.* Optical phased array technology. *Proc. IEEE* **84,** 268–298 (1996).
4. Meyer, R. A. Optical beam steering using a multichannel lithium tantalite crystal. *Appl. Opt.* **11,** 613–616 (1972).
5. Resler, D. P. High-efficiency liquid-crystal optical phased array beam steering. *Opt. Lett.* **21,** 689–691 (1996).
6. Ng, W. *et al.* The first demonstration of an optically steered microwave phased array antenna using true-time-delay. *IEEE J. Lightw. Technol.* **9,** 1124–1131 (1991).
7. Kwong, D., Hosseini, A., Zhang, Y. & Chen, R. T. 1 × 12 unequally spaced waveguide array for actively tuned optical phased array on a silicon nanomembrane. *Appl. Phys. Lett.* **99,** 051104 (2011).
8. Carlson, N. W. *et al.* Electronic beam steering in monolithic grating-surface-emitting diode laser arrays. *Appl. Phys. Lett.* **53,** 2275–2277 (1988).
9. Doylend, J. K. *et al.* Two-dimensional free-space beam steering with an optical phased array on silicon-on-insulator. *Opt. Express* **19,** 21595–21604 (2011).
10. Van Acoleyen, K., Rogier, H. & Baets, R. Two-dimensional optical phased array antenna on silicon-on-insulator. *Opt. Express* **18,** 13655–13660 (2010).
11. Van Acoleyen, K. *et al.* Off-chip beam steering with a one-dimensional optical pahsed array on silicon-on-insulator. *Opt. Lett.* **34,** 1477–1479 (2009).
12. Yu, N. *et al.* Light propagation with phase discontinuities: generalized laws of reflection and refraction. *Science* **334,** 333–337 (2011).
13. Yaacobi, A., Timurdogan, E. & Watts, M. R. Vertical emitting aperture nanoantennas. *Opt. Lett.* **37,** 1454–1456 (2012).
14. Abeles, J. H. & Deri, R. J. Suppression of sidelobes in the far-field radiation patterns of optical waveguide arrays. *Appl. Phys. Lett.* **53,** 1375–1377 (1988).
15. Mühlschlegel, P. *et al.* Resonant optical antennas. *Science* **308,** 1607–1609 (2005).
16. Schuck, P. J., Fromm, D. P., Sundaramurthy, A., Kino, G. S. & Moerner, W. E. Improving the mismatch between light and nanoscale objects with gold bowtie nanoantennas. *Phys. Rev. Lett.* **94,** 017402 (2005).
17. Kosako, T., Kadoya, Y. & Hofmann, H. F. Directional control of light by a nano-optical Yagi–Uda antenna. *Nature Photon.* **4,** 312–315 (2010).
18. Alù, A. & Engheta, N. Tuning the scattering response of optical nanoantennas with nanocircuit loads. *Nature Photon.* **2,** 307–310 (2008).
19. Tang, L. *et al.* Nanometre-scale germanium photodetector enhanced by a near-infrared dipole antenna. *Nature Photon.* **2,** 226–229 (2008).
20. Roelkens, G., Van Thourhout, D. & Baets, R. High efficiency silicon-on-insulator grating coupler based on a poly-silicon overlay. *Opt. Express* **14,** 11622–11630 (2006).
21. Taflove, A. & Hagness, S. C. *Computational Electrodynamics: The Finite-Difference Time-Domain Method* 3rd edn (Artech House, 2005).
22. Haupt, R. L. *Antenna Arrays: A Computational Approach* (Wiley, 2010).
23. Gerchberg, R. W. & Saxton, W. O. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik (Stuttg.)* **35,** 237–246 (1972).
24. Fienup, J. R. Reconstruction of an object from the modulus of its Fourier transform. *Opt. Lett.* **3,** 27–29 (1978).
25. Le Thomas, N. *et al.* Exploring light propagating in photonic crystals with Fourier optics. *J. Opt. Soc. Am. B* **24,** 2964–2971 (2007).
26. Padgett, M., Courtial, J. & Allen, L. Light's orbital angular momentum. *Phys. Today* **57,** 35–40 (2004).
27. Xia, C. *et al.* Supermodes for optical transmission. *Opt. Express* **19,** 16653–16664 (2011).
28. DeRose, C. T. *et al.* Low power and broadband 2×2 silicon thermo-optic switch. *Proc. Opt. Fiber Commun. Conf.* OThM3 (2011).
29. Watts, M. R. *et al.* Adiabatic thermo-optic Mach–Zehnder switch. *Opt. Lett..* (submitted).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Shtyrkova and J. S. Orcutt for help with the initial measurement setup, and also APIC Corporation. This work was supported by the Defense Advanced Research Projects Agency (DARPA) of the United States under the E-PHI and SWEEPER projects, grant no. HR0011-12-2-0007. J.S. acknowledges support from DARPA POEM award HR0011-11-C-0100.

Author Contributions J.S. and M.R.W. conceived the idea of the project. J.S. simulated and designed the devices and the phased array system, laid out the mask and performed the experimental characterizations and analysis. E.T. helped with the NPA system simulation algorithm. A.Y. contributed to the element antenna simulation. E.H. coordinated the mask. J.S. and M.R.W. wrote the paper. M.R.W. supervised the project. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R.W. (mwatts@mit.edu).

doi:10.1038/nature11731

Carbon-dioxide-rich silicate melt in the Earth's upper mantle

Rajdeep Dasgupta[1], Ananya Mallik[1], Kyusei Tsuno[1], Anthony C. Withers[2], Greg Hirth[3] & Marc M. Hirschmann[2]

The onset of melting in the Earth's upper mantle influences the thermal evolution of the planet, fluxes of key volatiles to the exosphere, and geochemical and geophysical properties of the mantle. Although carbonatitic melt could be stable 250 km or less beneath mid-oceanic ridges[1,2], owing to the small fraction (~0.03 wt%) its effects on the mantle properties are unclear. Geophysical measurements, however, suggest that melts of greater volume may be present at ~200 km (refs 3–5) but large melt fractions are thought to be restricted to shallower depths. Here we present experiments on carbonated peridotites over 2–5 GPa that constrain the location and the slope of the onset of silicate melting in the mantle. We find that the pressure–temperature slope of carbonated silicate melting is steeper than the solidus of volatile-free peridotite and that silicate melting of dry peridotite + CO_2 beneath ridges commences at ~180 km. Accounting for the effect of 50–200 p.p.m. H_2O on freezing point depression, the onset of silicate melting for a sub-ridge mantle with ~100 p.p.m. CO_2 becomes as deep as ~220–300 km. We suggest that, on a global scale, carbonated silicate melt generation at a redox front ~250–200 km deep[6], with destabilization of metal and majorite in the upwelling mantle, explains the oceanic low-velocity zone and the electrical conductivity structure of the mantle. In locally oxidized domains, deeper carbonated silicate melt may contribute to the seismic X-discontinuity. Furthermore, our results, along with the electrical conductivity of molten carbonated peridotite[7] and that of the oceanic upper mantle[5], suggest that mantle at depth is CO_2-rich but H_2O-poor. Finally, carbonated silicate melts restrict the stability of carbonatite in the Earth's deep upper mantle, and the inventory of carbon, H_2O and other highly incompatible elements at ridges becomes controlled by the flux of the former.

Knowledge of mid-oceanic-ridge basalt generation suggests that the initiation of silicate melting in an upwelling mantle is restricted to the top 70–85 km (refs 8, 9). This constraint is potentially at odds with observations of a seismic low-velocity zone that extends as deep as 200 km (ref. 4). Although the low shear-wave speed of oceanic mantle can perhaps be explained by high temperatures[10], or hydroxyl groups in nominally anhydrous minerals[11], correlation between velocity anomalies and high attenuation[3] may require the presence of partial melts at depths even in excess of 200 km. Intra-plate upper-mantle electrical conductivity, with a possible anomaly at ~220 km depth[5,12], may also require the presence of a highly conductive melt. Yet silicate melts are not thought to be generated at depths approaching or greater than 200 km.

If Earth's upper mantle is locally oxidized at depth, then carbonatitic melt could be generated atop the transition zone in upwelling mantle beneath ridges. However, with ≤100 p.p.m. CO_2 in the mantle, only ≤0.03 wt% of carbonatitic melt is produced and the effect of such a small melt fraction on mantle physical properties is unclear. Moreover, mantle at depths ≥250 km, on a global scale, is shown to be metal-saturated[6], thereby destabilizing a carbonatitic melt. Redox melting probably begins at 250–200 km depth in upwelling sub-ridge mantle, but the nature of generated melt at this redox front—that is, carbonate-rich versus silicate-rich—remains unknown. Dissolved $CO_2 \pm H_2O$ may stabilize a silicate melt at much lower temperatures than the volatile-free peridotite solidus, but previous work only constrained the stability of carbonated silicate melt in simplified systems ($CaO–MgO–Al_2O_3–SiO_2$ (CMAS) + CO_2 (see, for example, ref. 13) or $CaO–MgO–FeO–SiO_2 \pm Al_2O_3 + CO_2$ system (ref. 14)) or at a single pressure of 3 GPa (ref. 15) for a natural peridotite system. The key information that remained unconstrained was the location and the pressure–temperature P–T slope of the onset of carbonated silicate melting of natural fertile peridotite; that is, the slope of the melt CO_2 isopleths and carbonate–silicate melt transition in P–T space.

We conducted partial melting experiments with nominally anhydrous, carbonated fertile peridotite at temperatures between 1,300 and 1,750 °C and pressures between 2 and 5 GPa to constrain the pressure dependence of mixing between near-solidus carbonatite and higher-temperature silicate melts. Experiments were conducted with a piston-cylinder press (at a pressure of 2 GPa) and a Walker-style multi-anvil press (at pressures of 4 and 5 GPa) (Methods). A fertile peridotite with 2.5 wt% CO_2 (PERC) and 1 wt% CO_2 (PERC3) (Supplementary Table 1)[2] was used as the starting material.

Olivine (Ol), orthopyroxene (Opx) and carbonated partial melts (c-SL) are present in all the experiments in the presence or absence of garnet (Gt) and clinopyroxene (Cpx) (Fig. 1a and Supplementary Table 2). Near the solidus, the melt is carbonatitic (<10 wt% SiO_2 and 38–45 wt% CO_2 dissolved in the melt) with melt fractions F of ~5–6 wt% (PERC) or ~2–3 wt% (PERC3) that are directly proportional to the bulk CO_2 contents ($F = C^0_{CO2}/C^L_{CO2}$, assuming perfect incompatibility of carbon in residual silicates, where C^0_{CO2} and C^L_{CO2} are the concentrations of CO_2 in the mantle source and liquid, respectively). Carbonatitic melt evolves to a carbonated silicate melt (>25 wt% SiO_2 and ≤25 wt% CO_2 dissolved in the melt; Supplementary Table 3) with increasing temperature as the dissolution of Cpx and/or garnet into carbonatitic fluid becomes significant, melt fraction increases and C^L_{CO2} gets diluted.

The key observation of our experiments is that the temperature of onset of carbonated silicate melting, that is, stabilization of a melt with ~25 wt% CO_2, is significantly lower than the solidus of volatile-free peridotite[16] at the pressure range investigated. The transition from carbonatite to silicate melts in our study also occurs at a ~200 °C lower T than the similar transition observed in the simple CMAS + CO_2 system[13], suggesting a crucial function for TiO_2, Na_2O and FeO in stabilizing carbonated silicate melts. Our results also demonstrate that the temperature difference between the temperature–melt fraction T–F trends defined for the volatile-free peridotite and carbonated systems increases with increasing pressure between 2 and 5 GPa (Fig. 1b). Thus, at greater depths the onset of silicate melting (dissolution in a carbonatite) takes place at progressively lower temperatures with respect to the volatile-free peridotite solidus. We interpret the observations as indicating that this feature is caused by the diminishing influence of non-ideal mixing between the silicates and carbonates with increasing pressure; that is, increasing the separation of the

[1]Department of Earth Science, Rice University, 6100 Main Street, MS 126, Houston, Texas 77005, USA. [2]Department of Earth Sciences, University of Minnesota, 310 Pillsbury Drive, Minneapolis, Minnesota 55455, USA. [3]Department of Geological Sciences, Brown University, 324 Brook Street, Providence, Rhode Island 02912, USA.

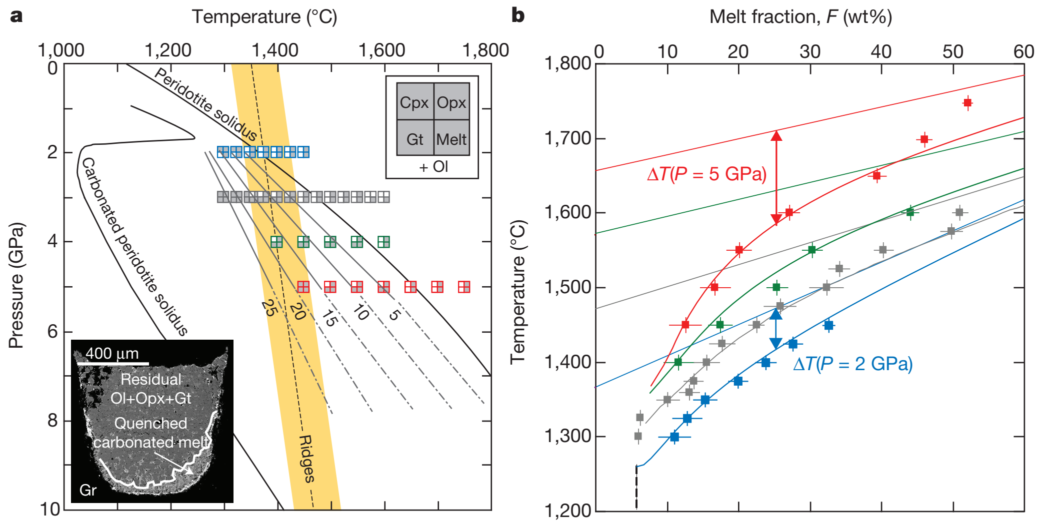


Figure 1 | Experimental conditions, phase assemblages, melt CO_2 isopleths and extent of melting as a function of temperature. **a**, Pressure–temperature plot. **b**, Temperature–melt-fraction plot. Results are shown from this (at 2 GPa (blue), 4 GPa (green) and 5 GPa (red)) and previous (at 3 GPa (grey)[15]) studies. The grey lines in **a** show the modelled melt CO_2 isopleths for carbonated silicate melts between the carbonated peridotite solidus[2,25] and the CO_2-free peridotite solidus[16]; numbers along the lines are percentages by weight of CO_2 in the melt. The segment of the carbonated peridotite solidus deeper than ~250 km is relevant only for locally oxidized domains of the mantle. The inset to **a** shows a BSE image of a typical experimental sample with residual Ol, Opx and Gt and quenched carbonated silicate melt contained in graphite (Gr) capsule. Also shown for comparison are the solid mantle adiabats relevant for oceanic ridges. In **b**, temperature–melt-fraction data from carbonated peridotite (curves with points; 2.5 wt% bulk CO_2; this study and ref. 15) are compared with similar data in CO_2-free systems (straight lines) from the literature (see Supplementary Information). For a given melt fraction, carbonated silicate melts are stabilized at a lower temperature than silicate melts from a volatile-free peridotite, and that temperature difference (for a constant CO_2 bulk composition) increases with increasing pressure from 2 to 5 GPa (marked by $\Delta T(P)$). The fit to the data for carbonated silicate melt fraction versus temperature is based on $\Delta T(P)$ versus melt CO_2 contents (Fig. 3; Supplementary Information). The melt CO_2 isopleths in **a** are based on $\Delta T(P)$ at each pressure from 2, 3, 4 and 5 GPa. For clarity, the data for 1wt% bulk CO_2 composition are not shown. Errors in **b** are 1σ with respect to the mean.

carbonated silicate melt liquidus from the metastable miscibility gap between carbonate and silicate melts (Supplementary Fig. 1). Indeed, the plot of carbonated silicate melt compositions (Supplementary Table 3), expressed in terms of molar $CO_2/(CO_2+SiO_2)$, as a function of pressure suggests that the transition from carbonate melt to silicate melt is sharp at lower pressures (for example, 2 GPa) but becomes more gradual with increasing pressure to 5 GPa (Fig. 2), which is consistent with observations in simplified and hydrous carbonated systems[14,17]. Owing to higher bulk CO_2, many of our peridotite partial melting experiments produce Cpx-free and some Gt-free assemblages. However, additional experiments at conditions appropriate for the carbonatite-silicate melt transition for peridotite (that is, 3 GPa and 1,375 °C) on more fertile bulk compositions (carbonated silicate melt fluxed peridotite) reveal that the mixing behaviour between carbonate-rich and silicate-rich melt depends little on peridotite bulk composition. Carbonate–silicate major-element mixing, as revealed by SiO_2–CO_2 and CaO–CO_2 systematics of partial melts in all our experiments, is similar to that for partial melts that are saturated with Ol + Opx + Cpx + Gt of variable compositions from this and previous studies (Supplementary Information and Supplementary Fig. 3).

Our experiments show that, in the presence of CO_2, appreciable melting or dissolution of mantle silicates at greater depths begins at temperatures significantly below the volatile-free solidus (Fig. 1b). However, Earth's mantle beneath oceanic ridges contains only 50–100 p.p.m. CO_2 (refs. 18–20), and the extent of CO_2-induced melting is proportional to source CO_2, $C^0_{CO_2}$ (ref. 15). Thus, to extrapolate our experimental data to natural mantle peridotite and to estimate the extent of silicate melting (F) induced by trace CO_2, we parameterize (following the approach of ref. 15) the temperature difference (ΔT) at a given melt fraction between the T–F trends for volatile-free peridotite (Supplementary Information) and for carbonated peridotite (this study, and ref. 15) as a function of CO_2 content of the experimental melts, $C^L_{CO_2}$ (Fig. 3). Figure 3 shows that at any given melt CO_2 content, ΔT increases with increasing P up to 5 GPa. If we assume that all the

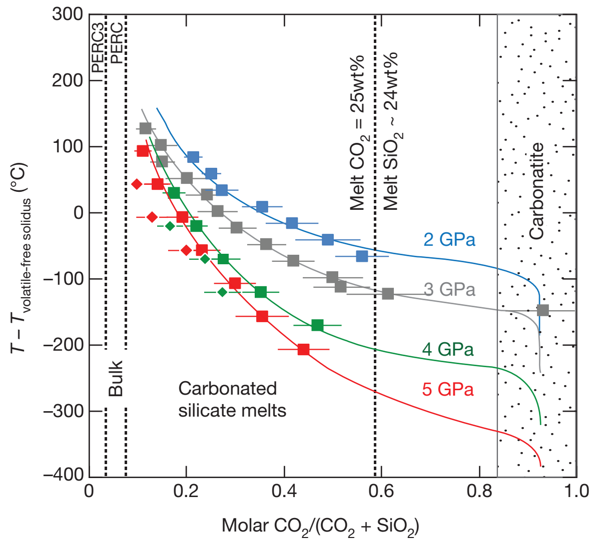


Figure 2 | Temperature–composition diagram showing evolution of melt composition. Melt compositions are expressed in terms of molar $CO_2/(CO_2 + SiO_2)$, as a function of reduced temperature ($T - T_{\text{volatile-free peridotite solidus}}$) for carbonated peridotite with 2.5 wt% (squares; this study and ref. 15) and 1 wt% CO_2 (diamonds; this study). The transition from carbonatite to carbonated silicate melt is sharp at lower pressures (≤3 GPa) but more gradual at higher pressures (≥4 GPa). The onset of carbonated silicate melting (arbitrarily placed at a molar $CO_2/(CO_2 + SiO_2)$ of ~0.6—that is, a melt CO_2 content of ~25 wt%)—occurs at a lower temperature with respect to the volatile-free peridotite solidus at higher pressures. Compositional errors are 1σ. Reduced temperature uncertainties are <25 °C and are not plotted.

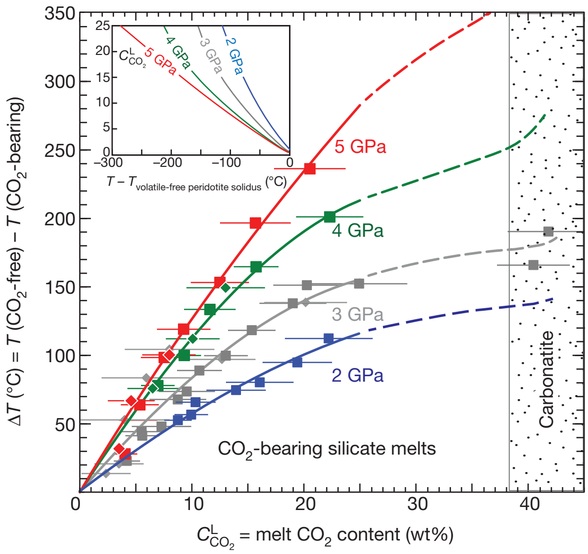


Figure 3 | Plot of ΔT as a function of concentration of CO_2 in the partial melts, C^L_{CO2}. ΔT is the difference between the T–F trends of CO_2-free peridotite and those of CO_2-bearing peridotite. Results from this study (squares, 2.5 wt% bulk CO_2; diamonds, 1 wt% bulk CO_2) and from ref. 15 are shown. At a given C^L_{CO2}, ΔT increases with increasing pressure from 2 to 5 GPa, suggesting a larger influence of CO_2 on silicate melting at greater depth. Inset: the modelled C^L_{CO2} as a function of reduced temperature (Fig. 2) at 2, 3, 4 and 5 GPa for p.p.m. level bulk CO_2. Errors are 1σ with respect to mean C^L_{CO2}.

carbon is dissolved in the melt, the C^L_{CO2}–$\Delta T(P)$ relations in Fig. 3 allow us to estimate $T - C^L_{CO2}$ and $T - F$ ($F = C^0_{CO2}/C^L_{CO2}$) trends at each pressure and for any given C^0_{CO2}. This analysis shows that trace carbonated silicate melting in the Earth's upper mantle begins at ~120 °C below the volatile-free peridotite solidus at 2 GPa, ~160 °C below the solidus at 3 GPa (ref. 15), ~210 °C below the solidus at 4 GPa, and ~280 °C below the solidus at 5 GPa.

We determined both the slopes of C^L_{CO2} isopleths in P–T space (Fig. 1) and the depth–melt fraction, depth–C^L_{CO2} trends for asthenospheric mantle with trace CO_2 (Fig. 4). Our model shows that a dry silicate melt with 25 wt% dissolved CO_2 is stabilized as deep as ~180 km for a sub-ridge mantle with a potential temperature T_P of 1,350 °C (145 km for 1,315 °C and 230 km for 1,400 °C T_P) and that 100 p.p.m. CO_2 alone can induce silicate melting with a melt fraction as much as ~0.07% at ~150 km depth beneath ridges. In contrast, previous estimates predicted a carbonatite melt fraction of only 0.02% at a similar depth[15]. Asthenospheric mantle, however, also contains trace quantities of an H_2O component; given the incompatible behaviour of H_2O during mantle melting[9,21], the presence of H_2O should deepen the onset of carbonated silicate melting[17], enhance the extent of melting at a given depth[15], and decrease C^L_{CO2}. To incorporate the effect of H_2O, we use the cryoscopic equation[22] with depth-dependent peridotite–melt partition coefficients for H_2O, D_{H2O} (ref. 9), by taking into account the possible effect of dissolved carbonate in the melt on D_{H2O} (Supplementary Information). Our calculations suggest that with 100 p.p.m. CO_2 and 50–200 p.p.m. H_2O in the sub-ridge mantle, 1–5 wt% H_2O can be dissolved into deep carbonated silicate melt with ~25 wt% CO_2 and that this concentration of dissolved H_2O will cause the onset of wet, carbonated melting to be as deep as 220–300 km. As much as 0.06–0.1 wt% wet carbonated silicate melt is generated by ~150 km depth, and 0.25 wt% by 100 km. For peridotite with 300 p.p.m. CO_2 and 300–500 p.p.m. H_2O, similar to that in the source of enriched

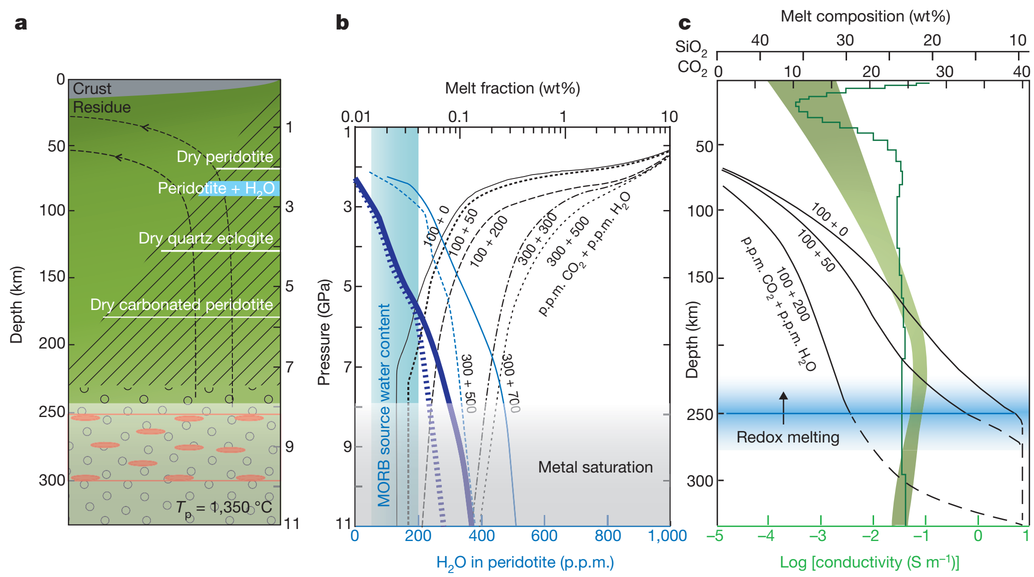


Figure 4 | Melting regime and mantle flow beneath a mid-oceanic ridge along a 1,350 °C mantle potential temperature (T_p) adiabat. Also shown are the extent of equilibrium melting, dehydration and melt composition of peridotite. **a**, The depth of onset of silicate melting (from stippled to hatched region) is relevant for a mantle with ~50 p.p.m. H_2O and 100 p.p.m. CO_2. Also shown are the silicate solidi of damp (50–200 p.p.m. H_2O) peridotite (blue band)[9], volatile-free peridotite[16], quartz eclogite (recycled oceanic crust)[31], and dry carbonated peridotite (this study). The transition from carbonatite to carbonated silicate melt is ≥110–150 km deeper than the solidus of volatile-free peridotite. The horizontal ellipses mark the depths of impedance contrast in oceanic mantle X-discontinuity[28] and the grey shaded region marks the stability of Ni-Fe metal[6]. **b**, Residual peridotite H_2O content (thin light blue, equilibrium melting; bold dark blue, fractional melting) is shown for two bulk compositions that are at their H_2O storage capacity (between 500 p.p.m. (dashed) and 700 p.p.m. (solid)) atop the transition zone and have 300 p.p.m. CO_2. The grey shaded region indicates that the curves are metastable owing to redox freezing of carbonate component with enhanced stability of Fe^{3+} in majorite and stability of Ni–Fe metal[1,6]. **c**, Electrical conductivity of the north Pacific[5] (green shaded region) and Philippine Sea mantle[26] (green curve) compared with the modelled C^L_{CO2} and C^L_{SiO2} for fertile peridotite with different CO_2–H_2O compositions. Solid segments are relevant where redox melting defines the solidus at ~250 km depth (blue shaded band). Dashed segments are relevant for oxidized domains. Because H_2O enhances melting and dilutes C^L_{CO2}, if highly conductive melt with >15–30 wt% CO_2 is required to produce the observed electrical conductivity anomaly at 180–250 km depth, the mantle needs to be H_2O-poor.

oceanic basalts[23,24], the equilibrium melt fraction can be as much as 1.5 wt% at 100 km, even for a mantle with a potential temperature similar to that below ridges (Fig. 4b).

The depth of onset of wet, carbonated silicate melting beneath ridges suggests that carbonatite melt stability in the oceanic upper mantle is restricted to $\gtrsim$220–300 km (Fig. 4c). However, the deep stability of carbonatite on a global scale is probably eliminated, because the depth of transition of carbonatite to carbonated silicate melt exceeds the depth of metal saturation and diamond stability[1,6] for mantle with >50 p.p.m. H_2O (Fig. 4c). Our results, along with the previous experiments on mantle redox state, suggest that the first-formed melt in the upwelling mantle beneath ridges, globally, is a carbonated silicate melt with a near-solidus melt fraction as much as 0.05–0.06 wt% at ~250–220 km (Fig. 4b). This melting commences at a redox front, where metal and majorite (the main host of Fe^{3+}) become unstable[6], promoting the stability of a carbonated silicate melt over diamond, carbide, carbon-bearing metal alloy or alloy melt[25]. Because mantle electrical conductivity profiles[5,26] require a melt with >15 wt% CO_2 (refs 7, 27), our data also suggest that the sub-ridge mantle at depths is CO_2-rich but H_2O-poor ($\leq$200 p.p.m.). Higher H_2O content in the mantle enhances melting and lowers the CO_2 content (Fig. 4c) and the conductivity of such melts. Redox melting at ~250 km can generate a carbonatitic melt (~38 wt% CO_2) for a dry mantle, and the melt CO_2 content remains high enough (>15 wt% CO_2) for a mantle with $\leq$200 p.p.m. H_2O to explain the peak and elevated electrical conductivity beneath oceans at ~220 and 100–250 km depth, respectively[5]. Thus, our study helps to reconcile the geophysical observations of the oceanic low-velocity zone down to ~200 km, the highly conductive asthenosphere, and the petrological models of mantle melting and redox (Fig. 4). Moreover, the depth for the onset of carbonated silicate melting may be similar to the depth of the X-discontinuity (270–330 km) in oceanic mantle[28], if the mantle is locally oxidized. We suggest that the appearance of hydrous carbonated silicate melt in oxidized mantle might explain the distinct impedance contrast across ~250–300 km that is observed for the seismic X-discontinuity. We note that the near-zero dihedral angle of the melt–mantle system at these depths implies that even a trace amount of melt will have large influence on the physical properties[29].

Our experiments and model further allow us to calculate the H_2O depletion of nominally anhydrous mantle as a function of depth beneath ridges (Fig. 4b) by using depth-dependent partition coefficients of H_2O between peridotite and basaltic melt[9] (Supplementary Information) and the trend in melt fraction against depth for sub-ridge mantle. These calculations suggest that, even for the extreme case of a mantle that is at its H_2O storage capacity atop the transition zone (perhaps resulting from hydrous melting across the 410-km discontinuity[30]), volatile depletion owing to deep carbonated silicate melt generation can result a sub-ridge residual mantle with $\leq$50 p.p.m. H_2O (Fig. 4b). However, the combined data on mantle electrical conductivity and volatile inventory of MORB suggest that deep mantle contains $\leq$200 p.p.m. H_2O. Finally, deep commencement of carbonated silicate melting in the Earth's upper mantle suggests that a kimberlitic melt with 20–30 wt% SiO_2 (Fig. 4c and Supplementary Fig. 3), rather than carbonatite, controls the flux of highly incompatible elements, H_2O and CO_2 at mid-oceanic ridges through mixing with the shallow depleted melts.

METHODS SUMMARY

Experiments at high pressure and temperature were performed with piston-cylinder and multi-anvil devices to simulate Earth's oceanic upper-mantle conditions (2–5 GPa; 1,300–1,750 °C). Starting compositions included natural mineral mixes equivalent to fertile mantle peridotite with 1.0 wt% and 2.5 wt% CO_2 and synthetic mixes with variable CO_2 contents. All starting mixes were dried to achieve nominally anhydrous conditions and reduced at an fO_2 lower than that of quartz–fayalite–magnetite buffer. All mixes were contained in graphite capsules with a platinum outer liner. Experimental products were analysed with a field-emission-gun scanning electron microscope and an electron probe microanalyser for texture and major-element compositions. To obtain representative melt compositions from a quenched melt pool comprising dendrites, a defocused electron beam was employed on serially polished surfaces. Mass balance calculations were used to estimate phase proportions in all experiments. Dissolved CO_2 contents of melts were estimated by electron microprobe deficits and also using mass balance calculations. To quantify the freezing-point depression and the enhanced melt production effect of trace CO_2 on silicate partial melting of peridotite, ΔT, the temperature difference between the T–F trend for volatile-free peridotite and that for carbonated peridotite (this study and ref. 15) was fitted as a function of dissolved CO_2 of the melt, C^{L}_{CO2}. The ΔT–C^{L}_{CO2} trends at each pressure were fitted with an empirical function (ref. 15). Using this parameterization, the C^{L}_{CO2} isopleths in pressure–temperature space were obtained. To obtain the combined effect of dissolved CO_2 and H_2O in stabilizing incipient silicate melt below the volatile-free peridotite solidus, the freezing-point depression effect of each species was added; that is, $\Delta T_{total} = \Delta T(C^{L}_{CO2}) + \Delta T(C^{L}_{H2O})$, where both C^{L}_{CO2} and C^{L}_{H2O} vary as a function of depth, the first owing to pressure-dependent mixing of silicate and carbonate melt components as constrained in this study, and the second owing to the pressure dependence of $D^{peridotite/melt}_{H2O}$ (ref. 9).

Full Methods and any associated references are available in the online version of the paper.

Received 15 August; accepted 31 October 2012.

1. Rohrbach, A. & Schmidt, M. W. Redox freezing and melting in the Earth's deep mantle resulting from carbon–iron redox coupling. *Nature* **472,** 209–212 (2011).
2. Dasgupta, R. & Hirschmann, M. M. Melting in the Earth's deep upper mantle caused by carbon dioxide. *Nature* **440,** 659–662 (2006).
3. Dalton, C. A., Ekström, G. & Dziewonski, A. M. The global attenuation structure of the upper mantle. *J. Geophys. Res.* **113,** B09303 (2008).
4. Hammond, W. C. & Toomey, D. R. Seismic velocity anisotropy and heterogeneity beneath the Mantle Electromagnetic and Tomography Experiment (MELT) region of the East Pacific Rise from analysis of P and S body waves. *J. Geophys. Res.* **108,** 2176 (2003).
5. Lizarralde, D., Chave, A., Hirth, G. & Schultz, A. Northeastern Pacific mantle conductivity profile from long-period magnetotelluric sounding using Hawaii-to-California submarine cable data. *J. Geophys. Res.* **100,** 17837–17854 (1995).
6. Rohrbach, A. *et al.* Metal saturation in the upper mantle. *Nature* **449,** 456–458 (2007).
7. Yoshino, T., McIsaac, E., Laumonier, M. & Katsura, T. Electrical conductivity of partial molten carbonate peridotite. *Phys. Earth Planet. Inter.* **194–195,** 1–9 (2012).
8. Langmuir, C. H., Klein, E. M. & Plank, T. in *Mantle Flow and Melt Generation at Mid-ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 183–280 (American Geophysical Union Monograph no. 71, 1992).
9. O'Leary, J. A., Gaetani, G. A. & Hauri, E. H. The effect of tetrahedral Al^{3+} on the partitioning of water between clinopyroxene and silicate melt. *Earth Planet. Sci. Lett.* **297,** 111–120 (2010).
10. Stixrude, L. & Lithgow-Bertelloni, C. Influence of phase transformations on lateral heterogeneity and dynamics in Earth's mantle. *Earth Planet. Sci. Lett.* **263,** 45–55 (2007).
11. Karato, S.-i. in *Inside the Subduction Factory* (ed. Eiler, J.) 135–152 (American Geophysical Union Monograph no. 138, 2003).
12. Baba, K., Chave, A. D., Evans, R. L., Hirth, G. & Mackie, R. L. Mantle dynamics beneath the East Pacific Rise at 17°S: insights from the Mantle Electromagnetic and Tomography (MELT) experiment. *J. Geophys. Res.* **111** (B2), B02101 (2006).
13. Gudfinnsson, G. & Presnall, D. C. Continuous gradations among primary carbonatitic, kimberlitic, melilititic, basaltic, picritic, and komatiitic melts in equilibrium with garnet lherzolite at 3–8 GPa. *J. Petrol.* **46,** 1645–1659 (2005).
14. Stagno, V. & Frost, D. J. Carbon speciation in the asthenosphere: Experimental measurements of the redox conditions at which carbonate-bearing melts coexist with graphite or diamond in peridotite assemblages. *Earth Planet. Sci. Lett.* **300,** 72–84 (2010).
15. Dasgupta, R., Hirschmann, M. M. & Smith, N. D. Water follows carbon: CO_2 incites deep silicate melting and dehydration beneath mid-ocean ridges. *Geology* **35,** 135–138 (2007).
16. Hirschmann, M. M. Mantle solidus: experimental constraints and the effects of peridotite composition. *Geochem. Geophys. Geosyst.* **1,** 1042 (2000).
17. Foley, S. F. *et al.* The composition of near-solidus melts of peridotite in the presence of CO_2 and H_2O between 40 and 60 kbar. *Lithos* **112** (Supplement 1), 274–283 (2009).
18. Cartigny, P., Pineau, F., Aubaud, C. & Javoy, M. Towards a consistent mantle carbon flux estimate: Insights from volatile systematics (H_2O/Ce, δD, CO_2/Nb) in the North Atlantic mantle (14° N and 34° N). *Earth Planet. Sci. Lett.* **265,** 672–685 (2008).
19. Marty, B. & Tolstikhin, I. N. CO_2 fluxes from mid-ocean ridges, arcs and plumes. *Chem. Geol.* **145,** 233–248 (1998).
20. Saal, A. E., Hauri, E., Langmuir, C. H. & Perfit, M. R. Vapour undersaturation in primitive mid-ocean-ridge basalt and the volatile content of Earth's upper mantle. *Nature* **419,** 451–455 (2002).
21. Michael, P. J. The concentration, behavior and storage of H_2O in the suboceanic upper mantle: implications for mantle metasomatism. *Geochim. Cosmochim. Acta* **52,** 555–566 (1988).

22. Hirschmann, M. M. Partial melt in the oceanic low velocity zone. *Phys. Earth Planet. Inter.* **179,** 60–71 (2010).
23. Aubaud, C., Pineau, F., Hekinian, R. & Javoy, M. Degassing of CO_2 and H_2O in submarine lavas from the Society hotspot. *Earth Planet. Sci. Lett.* **235,** 511–527 (2005).
24. Sobolev, A. V. & Chaussidon, M. H_2O concentrations in primary melts from supra-subduction zones and mid-ocean ridges: implications for H_2O storage and recycling in the mantle. *Earth Planet. Sci. Lett.* **137,** 45–55 (1996).
25. Dasgupta, R. & Hirschmann, M. M. The deep carbon cycle and melting in Earth's interior. *Earth Planet. Sci. Lett.* **298,** 1–13 (2010).
26. Baba, K. *et al.* Electrical conductivity imaging of the Philippine Sea upper mantle using seafloor magnetotelluric data. *Phys. Earth Planet. Inter.* **183,** 44–62 (2010).
27. Gaillard, F., Malki, M., Iacono-Marziano, G., Pichavant, M. & Scaillet, B. Carbonatite melts and electrical conductivity in the asthenosphere. *Science* **322,** 1363–1365 (2008).
28. Bagley, B. & Revenaugh, J. Upper mantle seismic shear discontinuities of the Pacific. *J. Geophys. Res.* **113** (B12), B12301 (2008).
29. Yoshino, T., Yamazaki, D. & Mibe, K. Well-wetted olivine grain boundaries in partially molten peridotite in the asthenosphere. *Earth Planet. Sci. Lett.* **283,** 167–173 (2009).
30. Bercovici, D. & Karato, S.-i. Whole-mantle convection and the transition-zone water filter. *Nature* **425,** 39–44 (2003).
31. Yasuda, A., Fujii, T. & Kurita, K. Melting phase relations of anhydrous mid-ocean ridge basalt from 3 to 20 GPa: implications for the behavior of subducted oceanic crust in the mantle. *J. Geophys. Res.* **99,** 9401–9414 (1994).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Ballhaus for reviewing the manuscript. This study received support from the National Science Foundation and a Packard fellowship to R.D.

Author Contributions R.D. designed the project, performed most of the experiments and sample analyses, and wrote the paper. A.M. conducted the experiments and analyses of carbonated silicate melt fluxed peridotite compositions. K.T. participated in conducting the piston cylinder experiments and analyses of peridotite + CO_2 bulk compositions. A.C.W. participated in the multi-anvil experiments. All authors, including G.H. and M.M.H., participated in the discussion and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.D. (Rajdeep.Dasgupta@rice.edu).

METHODS

The starting compositions for peridotite + CO_2 partial melting experiments (PERC and PERC3: MixKLB-1 + mixture of carbonates) with 2.5wt% and 1.0wt% bulk CO_2 were the same materials used in refs 2 and 15 (Supplementary Table 1). This ensured reliable comparison between present and previous data (ref. 15). The base silicate powder was dried and reduced at f_{O_2} lower than that buffered by quartz–fayalite–magnetite (QFM-1). The mix was then reground under ethanol and CO_2 was added as a mixture of natural and reagent grade ($CaCO_3$, Na_2CO_3, K_2CO_3) carbonates in a manner that kept the Ca:Mg:Fe:Na:K ratio of the base peridotite (MixKLB-1) unaffected by CO_2 addition. This approach of CO_2 addition has been demonstrated to minimize the effect of overabundances of carbonates/CO_2 on the melting-phase relations of peridotite[32]. To test the nature of carbonate–silicate mixing, as captured by the compositions of carbonated silicate melts, as a function of variable peridotite compositions, additional experiments were performed on more fertile bulk compositions (Supplementary Table 1). These additional starting compositions were prepared with reagent-grade oxides and carbonates. All the carbonate compounds were fired in a 1-atm furnace at 110–300 °C for 4–12 h before mixing them with silicates under ethanol. Homogeneous mixtures of carbonate-bearing peridotite mix were loaded in graphite crucibles with platinum outer capsules (3.5 mm outer diameter for piston-cylinder (PC) experiments and 2.0 mm outer diameter for multi-anvil (MA) experiments). The loaded capsules were placed in a drying oven at 110 °C for a couple of hours before welding. PC experiments were conducted at Rice University, and MA experiments were conducted at the University of Minnesota. The PC and MA experiments followed the procedures, assemblies and P–T calibrations described and reported in detail in refs 33 and 34, respectively. Pressure uncertainties are believed to be ±0.1 GPa for PC and ±0.3 GPa for MA experiments, and the temperature uncertainties are ±10 °C for both the devices. The redox states during the experiments were not controlled strictly; however, the association of orthopyroxene, carbonated melt, olivine, clinopyroxene and graphite/diamond (inner capsule) allowed calculation of oxygen fugacity following the buffer reactions EDDOG/EDDOD (reaction 1) and/or EMOG/EMOD (reaction 2):

$$2Mg_2Si_2O_6 + CaMg(CO_3)_2 = CaMgSi_2O_6 + 2Mg_2SiO_4 + 2C + O_2 \quad (1)$$

$$Mg_2Si_2O_6 + 2MgCO_3 = 2Mg_2SiO_4 + 2C + 2O_2 \quad (2)$$

Estimated oxygen fugacities based on a recent calibration[14], which takes into account the effect of dilution of CO_2 in the melts on the equilibrium oxygen fugacity, are reported in Supplementary Table 2.

Run products were polished on soft nylon and velvet cloth by using dry polycrystalline diamond powders to achieve a 0.3 μm finish. H_2O and other liquid lubricants were avoided during polishing to minimize damage to the quenched carbonate-bearing melt. Textural and wavelength dispersive spectroscopic (WDS) phase compositional analysis of the carbon-coated experimental samples was performed with a Cameca SX-100 electron probe at NASA Johnson Space Center (peridotite + CO_2 experiments) and a Cameca SX-50 electron probe at Texas A&M University (carbonated silicate melt fluxed peridotite experiments). Analytical standards included a synthetic basalt glass (Ca, Mg, Fe, Si), an Indian Ocean basalt (Ca, Al, Si, Mg), olivine (Mg, Si, Fe), calcite (Ca), dolomite (Mg, Ca), Cpx (Ca, Mg, Fe, Si), enstatite (Mg, Si), pyrope (Mg, Al, Si), andradite (Ca, Al, Fe), oligoclase/albite (Na, Al), rutile/ilmenite (Ti), rhodonite (Mn), spessartine (Mn), potassium feldspar (Al, K) and chromite (Cr). To obtain meaningful average melt compositions from heterogeneously quenched dendritic melt pool texture, a defocused (5–30-μm spot size) electron beam was employed. Moreover, to obtain as many spot analyses as possible, two to four new longitudinal surfaces of each sample were exposed by serial polishing. To avoid sampling melt compositions changed by quench modification, only the melt regions that are separated from the residual mineral assemblage domains (Fig. 1) were analysed. Dissolved CO_2 concentrations of partial melts were estimated both by the difference between electron probe totals and 100wt% and by mass balance calculations (Supplementary Table 3). Because the two approaches yield similar values (Supplementary Fig. 4), those obtained from electron microprobe totals are plotted in all figures. Supplementary Table 2 reports the experimental conditions, obtained phase assemblages and phase proportions obtained by mass balance calculation. Supplementary Table 3 reports the major-element compositions of all resulting phases.

Freezing-point depression and the enhanced melt production effect of CO_2 on partial melting of peridotitic silicates was calculated by fitting ΔT, the temperature difference between the T–F trend for volatile-free peridotite and that for carbonated peridotite (this study, and ref. 15) as a function of dissolved CO_2 of the melt, C^{L}_{CO2}. The ΔT–C^{L}_{CO2} trends at each pressure are nearly linear up to a C^{L}_{CO2} of ~25wt% and were fitted with an empirical function,

$$\Delta T(^{\circ}C) = a \times C^{L}_{CO2} + b \times \ln[(100 - c \times C^{L}_{CO2})/100]$$

where C^{L}_{CO2} is in percentage by weight and a, b and c are constants that vary as a function of pressure (Supplementary Information).

To obtain the combined effect of dissolved CO_2 and H_2O in stabilizing minute amounts of silicate melt below the volatile-free peridotite solidus, we combined the freezing-point depression effect of each species; that is, $\Delta T_{total} = \Delta T(C^{L}_{CO2}) + \Delta T(C^{L}_{H2O})$, where both C^{L}_{CO2} and C^{L}_{H2O} vary as a function of depth, the first owing to pressure-dependent mixing of silicate and carbonate melt components as constrained in this study and the second owing to the pressure dependence of $D^{peridotite/melt}_{H2O}$ (ref. 9).

32. Dasgupta, R. & Hirschmann, M. M. Effect of variable carbonate concentration on the solidus of mantle peridotite. *Am. Mineral.* **92,** 370–379 (2007).
33. Tsuno, K. & Dasgupta, R. Melting phase relation of nominally anhydrous, carbonated pelitic-eclogite at 2.5–3.0 GPa and deep cycling of sedimentary carbon. *Contrib. Mineral. Petrol.* **161,** 743–763 (2011).
34. Dasgupta, R., Hirschmann, M. M. & Withers, A. C. Deep global cycling of carbon constrained by the solidus of anhydrous, carbonated eclogite under upper mantle conditions. *Earth Planet. Sci. Lett.* **227,** 73–85 (2004).

doi:10.1038/nature11744

Crystallographic snapshot of cellulose synthesis and membrane translocation

Jacob L. W. Morgan[1], Joanna Strumillo[1,2] & Jochen Zimmer[1]

Cellulose, the most abundant biological macromolecule, is an extracellular, linear polymer of glucose molecules. It represents an essential component of plant cell walls but is also found in algae and bacteria. In bacteria, cellulose production frequently correlates with the formation of biofilms, a sessile, multicellular growth form. Cellulose synthesis and transport across the inner bacterial membrane is mediated by a complex of the membrane-integrated catalytic BcsA subunit and the membrane-anchored, periplasmic BcsB protein. Here we present the crystal structure of a complex of BcsA and BcsB from *Rhodobacter sphaeroides* containing a translocating polysaccharide. The structure of the BcsA–BcsB translocation intermediate reveals the architecture of the cellulose synthase, demonstrates how BcsA forms a cellulose-conducting channel, and suggests a model for the coupling of cellulose synthesis and translocation in which the nascent polysaccharide is extended by one glucose molecule at a time.

Polysaccharides perform multiple vital functions in life, as reflected in their great variety in chemical composition and physicochemical properties. Extracellular polysaccharides, such as cellulose, chitin and hyaluronan, primarily perform structural functions and are synthesized inside the cell from nucleotide-activated donor sugars[1–3].

Cellulose is the most abundant biological polymer and consists of glucose molecules that are connected between their C1 and C4 carbons via acetal linkages[4]. It is predominantly generated by vascular plants and a large number of algae, but also by some bacteria[5,6], protists[7] and tunicates[8]. Cellulose synthases (CESAs) are membrane-embedded glycosyltransferases, which use UDP-activated glucose (UDP-Glc) to elongate the nascent polysaccharide processively in a reaction that inverts the configuration at the anomeric carbon of the newly added sugar from α to β[6,9,10]. Prokaryotic and eukaryotic CESAs share a similar predicted topology including eight transmembrane helices and at least one extended intracellular glycosyltransferase (GT) loop (Supplementary Fig. 1).

Bacteria, primarily Gram-negatives, produce and secrete cellulose via a protein complex consisting of at least three subunits (BcsA, BcsB and BcsC)[11]. The inner membrane protein BcsA is the catalytically active subunit and contains a conserved family-2 GT domain between transmembrane helices 4 and 5 (TM4 and TM5)[12]. BcsB is a periplasmic protein that is anchored to the inner membrane via a single, carboxy-terminal transmembrane helix. BcsA and BcsB are fused as a single polypeptide in some species, supporting the genetic observation that BcsB is essential for cellulose synthesis[13,14]. BcsC, required for cellulose synthesis *in vivo* but not *in vitro*[14], is predicted to form an 18-stranded β-barrel in the outer membrane. Some bacteria contain additional, non-essential, periplasmic subunits of unknown function[11,15].

Although some bacteria constitutively produce cellulose[6], bacterial cellulose production is often concomitant with the formation of sessile, multicellular aggregates, termed biofilms[11,16]. Biofilm bacteria are encased in a network of polysaccharides and proteinaceous fibres and are of particular concern to human health due to an increased tolerance to antimicrobial treatments[17]. Biofilm formation is stimulated by the bacterial secondary messenger cyclic-di-GMP, which bacteria sense via PilZ domains[18]. Accordingly, BcsA contains a PilZ domain within its C-terminal, intracellular domain and its activity is strongly stimulated by cyclic-di-GMP[19,20].

In all organisms, the cellulose polymer has to be exported across the plasma membrane. However, the mechanism by which a polymer consisting of thousands of glucose units can be translocated has been unclear. To understand this process, we determined the crystal structure of the catalytically active BcsA–BcsB complex in an intermediate state during cellulose synthesis and translocation at 3.25 Å resolution (Supplementary Table 1).

Architecture of the BcsA–BcsB complex

BcsA and BcsB form a 1:1 stoichiometric complex spanning approximately 150 Å perpendicular and 55 Å parallel to the membrane. The complex is divided into a cuboid-shaped membrane-spanning region sandwiched between large cytoplasmic and periplasmic domains (Fig. 1a). The BcsA–BcsB interface is approximately 4,500 $Å^2$ including a cluster of negative and positive charges on BcsA and BcsB, respectively (Supplementary Fig. 2).

BcsA contains four amino-terminal and four carboxy-terminal transmembrane helices separated by a large intracellular loop (4/5 loop) that forms a GT domain (amino acids 128–368) (Fig. 1a and Supplementary Fig. 3). TM3–8 form a narrow channel for the translocating polysaccharide, and BcsA's intracellular C terminus (amino acids 575–759) contains a six-stranded β-barrel and a highly curved α-helical region that attaches the β-barrel to the GT domain (Fig. 1a).

BcsB is a dome-shaped, β-strand-rich, periplasmic protein. Its N-terminal region forms the tip of the dome, whereas the C-terminal transmembrane anchor interacts with BcsA (Fig. 1a). Two amphipathic helices further stabilize its interaction with BcsA and the periplasmic water–membrane interface (Fig. 1a).

Throughout model building, we noticed a continuous, strong positive difference Fourier electron density that could not be accounted for by BcsA or BcsB (Fig. 1b). The ribbon-shaped electron density extended from BcsA's GT domain in a 45° angle towards the membrane, kinked near the putative water–lipid interface to run through the centre of the transmembrane region, and bent sharply at the BcsA–BcsB interface to emerge from the complex at approximately

[1]Center for Membrane Biology, Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908, USA. [2]Faculty of Biology and Environmental Protection, University of Lodz, Pilarskiego Street 14/16, 90-231 Lodz, Poland.

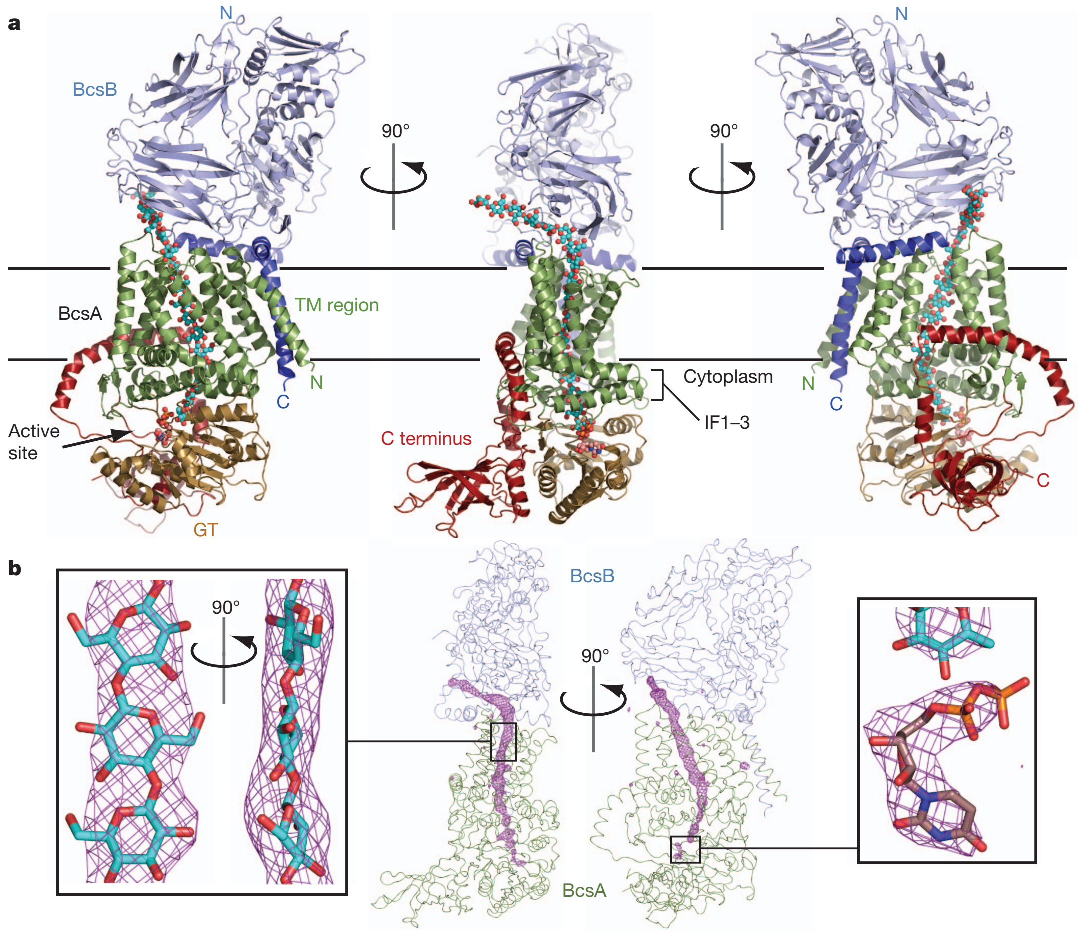


Figure 1 | Architecture of the BcsA–BcsB complex. a, BcsA and BcsB form an elongated complex with large cytosolic and periplasmic domains. The transmembrane helices of BcsA are coloured green, the glycosyltransferase (GT) domain beige, and the C-terminal domain red. BcsB is shown in light and dark blue for its periplasmic and membrane-associated regions, respectively. The N and C termini of both subunits are indicated and the translocating glucan and UDP are shown as cyan and violet spheres. Horizontal bars indicate the membrane boundaries. IF, amphipathic interface helices of BcsA. **b**, Unbiased $F_o - F_c$ difference Fourier electron density (pink mesh, contoured at 4.5σ) calculated before modelling the glucan and UDP molecules. The continuous density runs from the intracellular catalytic site to the periplasmic BcsA–BcsB interface and accommodates 18 glucose molecules (cyan sticks). The UDP density was contoured at 3σ (right panel).

145° (Fig. 1b). Past the BcsA–BcsB interface, the electron density extended into the bulk solvent. The characteristic shape of the electron density, its proximity to conserved residues throughout its path (see below), its origin at the glycosyltransferase catalytic site, and extension past the BcsA–BcsB complex strongly suggest that it represents a translocating glucan probably synthesized during or before crystallization.

The catalytic activity of the purified, detergent-solubilized BcsA–BcsB complex was confirmed by incubating it with UDP-Glc and cyclic-di-GMP. The synthesized polymer, that could be completely degraded by β-1,4-glucanase (data not shown), was sedimented, purified and analysed by glycosyl linkage analysis (performed by the Complex Carbohydrate Research Center, University of Georgia), confirming that 94% of it consists of 4-linked glucose (Supplementary Fig. 4).

Although probably reflecting an average of different registers of the translocating glucan, the density discerns the overall position of 18 glucose molecules (Fig. 1b and Supplementary Fig. 5). According to the prevailing model that cellulose synthase elongates the non-reducing end of the growing polymer[21,22], the glucan was modelled with its non-reducing end (glucose-18) at the glycosyltransferase catalytic site. At the active site, we observed additional, weak electron density detached from the glucan. On the basis of its shape and coordination by residues that are predicted to be implicated in nucleotide binding and glycosyl transfer[23], it seemingly represents a weakly bound UDP molecule (Fig. 1b and Supplementary Fig. 5). Therefore, our structure probably represents a cellulose synthase–cellulose translocation intermediate captured after glycosyl transfer and before replacing UDP with UDP-Glc.

The glycosyltransferase domain

The GT domain of BcsA contains the conserved signature D,D,D,Q(Q/R)XRW of three variably spaced aspartic acids and a pentapeptide consisting of a Gln that is often followed by a Gln or Arg, a variable residue, and an Arg and Trp. This signature is shared with other processive β-glycosyltransferases, such as chitin and hyaluronan synthases[2,24].

The GT domain adopts a glycosyltransferase A fold (GT-A)[10] consisting of a mixed, seven-stranded β-sheet that is surrounded by seven α-helices and attaches to the transmembrane region via three amphipathic interface (IF) helices (IF1–3) (Figs 1–3 and Supplementary Fig. 3). IF1 is formed by residues belonging to the 4/5 loop (amino acids 298–312) and packs against the C termini of BcsA's TM8 and TM4 as well as the N terminus of TM5. The second helix, IF2 (amino acids 373–395), directly precedes TM5 and includes the invariant Q(Q/R)XRW sequence of the signature. IF3 (amino acids 470–498) follows TM6 and forms a crescent-shaped helix around the cytosolic entrance to the transmembrane channel (discussed below). It is connected to TM7 via a conserved loop (IF3–TM7 loop) that runs across the entrance to the GT substrate-binding site (Fig. 2b and Supplementary Fig. 3). The GT domain protrudes from these IF helices

at approximately 45° (Fig. 1a). The inverting glycosyltransferase SpsA from *Bacillus subtilis*[23] superimposes with BcsA's GT domain with 2.15 Å root mean square deviation (r.m.s.d.) between Cα atoms (Supplementary Fig. 6).

The structure of the cellulose synthase reveals the function of the signature D,D,D,Q(Q/R)XRW as well as other highly conserved sequence motifs. Consistent with non-processive glycosyltransferases[23,25,26], the first two aspartic acids of the signature, Asp 179 and Asp 246, coordinate UDP (Fig. 2a). The third 'D' (Asp 343), which is part of the invariant 'TED' motif, probably represents the catalytic base owing to its proximity to the non-reducing end of the glucan (Fig. 2a).

The Q(Q/R)XRW part of the signature, together with an equally conserved FFCGS sequence, forms a binding site for the terminal disaccharide of the glucan, the acceptor (Fig. 2a). The Q(Q/R)XRW sequence belonging to IF2 is a part of the cytoplasmic entry to the glucan channel (Fig. 3). The first Q, Gln 379, forms a hydrogen bond to the guanidino group of the neighbouring Arg 380, which in turn forms a salt bridge to the invariant Glu 297. Arg 382, preceding the conserved Trp residue, most likely coordinates the diphosphate of UDP; however, the electron density for its side chain is weak, consistent with the apparent low occupancy of UDP. Trp 383 forms van der Waals interactions with the penultimate glucose molecule at the acceptor site (glucose-17), and packs against the Gln 379-Arg 380 pair towards the distal site of the catalytic pocket (Fig. 2a). The FFCGS pentapeptide (amino acids 316–320) interacts with the opposite side of the terminal disaccharide, primarily via the backbone carbonyls of Cys 318 and Gly 319 with glucose-18.

The dominant-negative *thanatos* mutation (Pro578Ser) in *Arabidopsis thaliana* CESA3[27] maps to the invariant QTPH sequence in bacterial cellulose synthases (Supplementary Fig. 1). His 276 of this motif is within hydrogen-bonding distance to glucose-17 (Fig. 2a) and is thus probably also involved in positioning the non-reducing end with respect to the catalytic base. The motif is part of an extended β-strand and Pro 275 (equivalent to Pro 578 in *A. thaliana* CESA3) induces a 'kink' in the β-strand that positions the side chain of His 276 towards the glucan, which might be disrupted when Pro is replaced by another amino acid.

A cavity next to the UDP-binding pocket might accommodate the glucose donor in the UDP-Glc-bound state. This pocket is flanked on one side by Ala 225 and Lys 226 of the invariant AKAGN motif and, on the other side, by Glu 342 (of the conserved TED motif) and His 224 (Fig. 2a and Supplementary Fig. 7).

Activation by cyclic-di-GMP

The activity of BcsA is stimulated by cyclic-di-GMP[20]. The cyclic-di-GMP-responsive PilZ domain is localized within BcsA's C terminus[18], right next to the GT domain (Fig. 2b). The C terminus of BcsA extends from the last transmembrane helix (TM8) via a short β-strand, which forms a two-stranded β-sheet with the IF3–TM7 loop, before it folds into a six-stranded β-barrel (amino acids 585–675). The β-barrel points away from the GT domain at approximately 90°. Past the β-barrel, the polypeptide chain continues in an α-helical conformation on the surface of the GT domain towards the water–lipid interface. At the interface, it forms an amphipathic helix that interacts with TM6, IF3 and TM8 (Fig. 1a) before it curves away from the membrane towards the GT domain and breaks at Glu 743. After Glu 743, the polypeptide continues in an extended conformation towards the β-barrel, past the substrate-binding pocket.

The β-barrel of BcsA aligns with an r.m.s.d. of 2.2 Å between Cα atoms with the cyclic-di-GMP binding protein VCA0042[28] (Supplementary Fig. 8). The conserved residues Arg 584, Asp 609, Ser 611 and Gly 614 on the β-barrel surface as well as Gln 578 and Arg 579 of the TM8–β-barrel linker are probably implicated in cyclic-di-GMP binding[28] (Fig. 2b and Supplementary Fig. 8). We hypothesize that cyclic-di-GMP binding to BcsA induces conformational changes that allow UDP-Glc to access the catalytic site. A likely candidate for displacement is the IF3–TM7 loop that runs across the entrance to the catalytic pocket and interacts with the TM8–β-barrel linker (Fig. 2b and Supplementary Fig. 3). Notably, replacing Thr

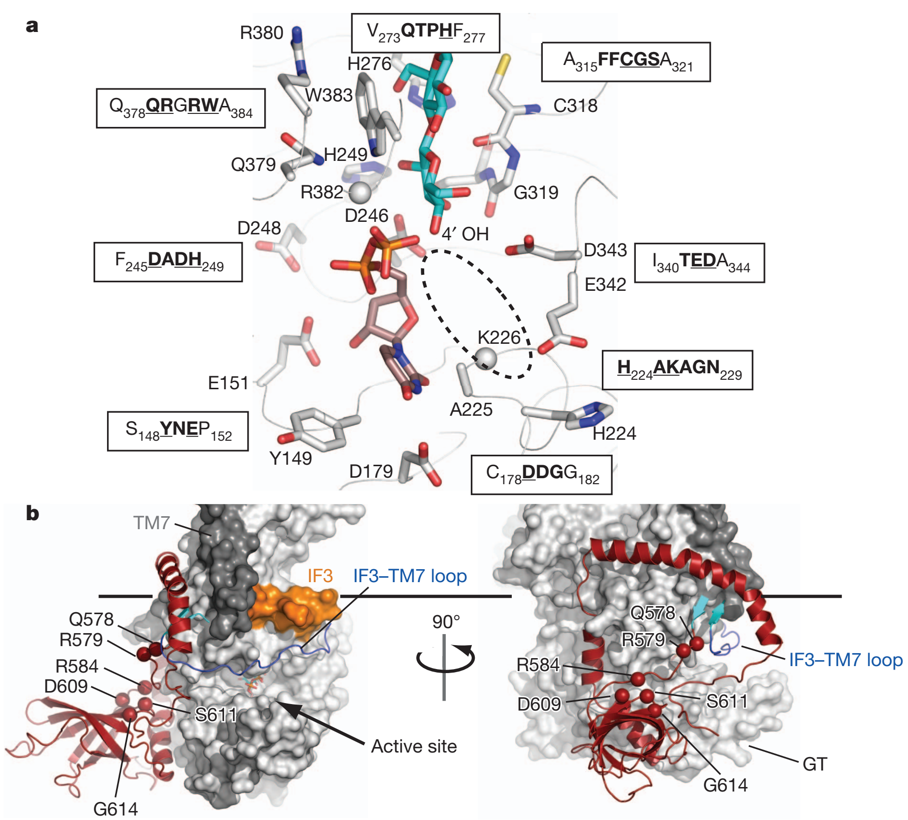


Figure 2 | Organization of BcsA's catalytic site and PilZ domain. a, Conserved residues of BcsA coordinate UDP and the terminal disaccharide of the glucan. Side chains represented in sticks belong to the sequence motifs shown in single letter code. Conserved residues are highlighted in bold and the depicted residues are underlined. All side chains shown could be positioned unambiguously in the electron density map. No density was observed for the side chains of Lys 226 and Arg 382 (shown as spheres for their Cα atoms). The probable position of the donor Glc is indicated by a dashed ellipsoid. **b**, Surface representation of BcsA with its C-terminal domain shown as cartoon in red. The linker connecting TM8 with the β-barrel forms a two-stranded β-sheet (shown in cyan) with the IF3–TM7 loop. TM7 is coloured dark grey and IF3 is orange. The Cα atoms of residues likely involved in cyclic-di-GMP binding are shown as spheres and are labelled. The horizontal bar indicates the cytoplasmic membrane boundary.

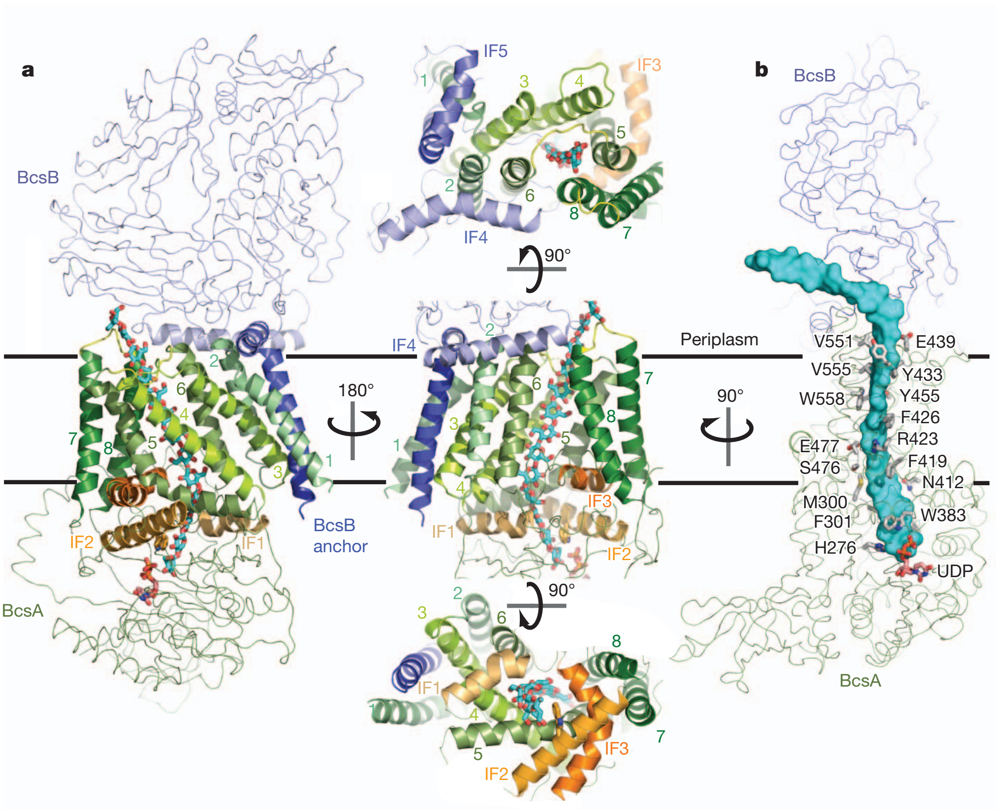


Figure 3 | The membrane-integrated domain of BcsA–BcsB. **a**, The transmembrane region includes nine transmembrane helices and three cytoplasmic and two periplasmic interface helices (IF1–5). BcsA and BcsB are shown as green and blue ribbons, respectively. Transmembrane and interface helices are show in cartoon representation and coloured in shades of green and orange for BcsA's TM1–8 and IF1–3, respectively, and shades of blue for BcsB's IF4 and IF5 and its C-terminal transmembrane anchor. TM3–8 of BcsA form a narrow channel. Its cytoplasmic entrance is formed by IF1–3 as well as the N-terminal half of TM5 and its periplasmic exit is between the periplasmic 5/6 and 7/8 loops of BcsA. Trp 383 of the Q(Q/R)XRW motif is shown in stick representation. **b**, Conserved residues of BcsA interact with the translocating glucan. BcsA and BcsB are shown as ribbons, and side chains of BcsA contacting the glucan (cyan surface) are shown as sticks. Residues of BcsB contacting the polymer are not conserved and are not shown.

942 with Ile in *A. thaliana* CESA3 confers resistance to the herbicide isoxaben[29]. The corresponding, equally conserved residue in our structure is Thr 506 in the IF3–TM7 loop (Supplementary Fig. 1), supporting the hypothesis that this loop is involved in regulating the activity of BcsA.

The transmembrane region and polysaccharide channel

The transmembrane region of the cellulose synthase includes eight transmembrane helices from BcsA (TM1–8) and one from BcsB (Fig. 3a). The helices are framed by three cytoplasmic and two periplasmic IF-helices formed by BcsA and BcsB, respectively. In addition, the N-terminal four turns of TM5 also run parallel to the membrane before the helix bends towards the hydrocarbon core at Pro 420 (Fig. 3a). All transmembrane helices, except TM6 and TM7 and BcsB's transmembrane helix, tilt with respect to the membrane normal and surround a narrow channel approximately 8 Å wide and 33 Å long. Because the transmembrane helices are organized in pairs in one direction (helices 1/2, 3/6, 4/8 and 5/7 of BcsA), their overall arrangement resembles an elongated cuboid, and BcsB's transmembrane anchor packs against one side of the cuboid (Fig. 3a). On the membrane periplasmic side, BcsB forms two amphipathic helices (IF4 and IF5) at the water–lipid interface. IF4 (amino acids 569 to 592) interacts with the periplasmic termini of BcsA's TM2, TM6 and TM8, and IF5 (amino acids 679–693) directly precedes BcsB's membrane anchor (amino acids 695–720) and contacts BcsA's TM1 and the 1/2 loop (Fig. 3a).

The channel formed by BcsA's TM3–8 accommodates ten glucose units of the translocating glucan (Fig. 3). The glucan enters the channel through the cytoplasmic opening formed by IF1–3 and the N-terminal half of TM5, crosses the membrane parallel to TM5 and TM6, and exits on the periplasmic side between BcsA's 5/6 and 7/8 loops at the BcsA–BcsB interface. Here, the glucan kinks to protrude from the complex sideways (Fig. 3b). TM6 and TM8 are separated by about 9 Å and 15 Å (between Cα atoms) at the periplasmic and cytoplasmic sides of the membrane, respectively, and only the tips of their side chains maintain the barrier towards the lipids (Fig. 3a). In the absence of a translocating glucan, transmembrane helices 7 and 8 might move towards the N-terminal transmembrane helices to close the channel.

All transmembrane helices of BcsA, with the exception of TM1 and TM2, directly contact the glucan. The faces of the glucopyranose rings form van der Waals interactions with residues Met 300 and Phe 301 of IF1, Phe 316 directly proceeding IF1, Trp 383 of IF2, Phe 419, Phe 426 and Tyr 433 of TM5, Phe 441 in the 5/6 loop, and Val 551, Val 555 and Trp 558 of TM8 (Fig. 3b). The equatorial hydroxyl groups of the glucose units form hydrophilic contacts with Tyr 80 of TM3, Asn 118 of TM4, His 276 of the GT domain, Asn 412 and Arg 423 of TM5, Glu 439 in the TM5/6 loop, Tyr 455 of TM6, and Ser 476 and Glu 477 of IF3. We note that the quinoxyphen-resistance mutation Ala903Val in *A. thaliana* CESA1[30] aligns with Tyr 455 in TM6, which forms a hydrogen bond to the translocating glucan (Fig. 3b and Supplementary Fig. 1). Therefore, although the sequence alignment in this region is weak, it is likely that quinoxyphen affects the translocation rather than the synthesis of the glucan.

The periplasmic domain

BcsB protrudes approximately 60 Å into the periplasm with its N-terminal half (amino acids 54–307) forming the membrane distal part. BcsB consists primarily of β-strands that are organized into distinct domains, including two jellyroll and two flavodoxin-like folds (Fig. 4). The jellyrolls (amino acids 54–185 and 316–452) show a striking structural similarity with carbohydrate binding modules (Supplementary Fig. 9)[31] and are termed carbohydrate-binding domains (CBD) 1 and 2. Both CBDs contain eight antiparallel β-strands separated by loops 2–16 residues in length and adopt a similar fold, such that their Cα atoms superimpose with an r.m.s.d. of 2.5 Å (Supplementary Fig. 9). The CBDs are positioned at a 45° angle relative to one another and interact via two extended loops, including a disulphide bond between the conserved Cys 163 and Cys 430 (Fig. 4a).

CBD2 runs parallel to the plane of the membrane above the periplasmic exit of the transmembrane channel. It interacts with BcsA's

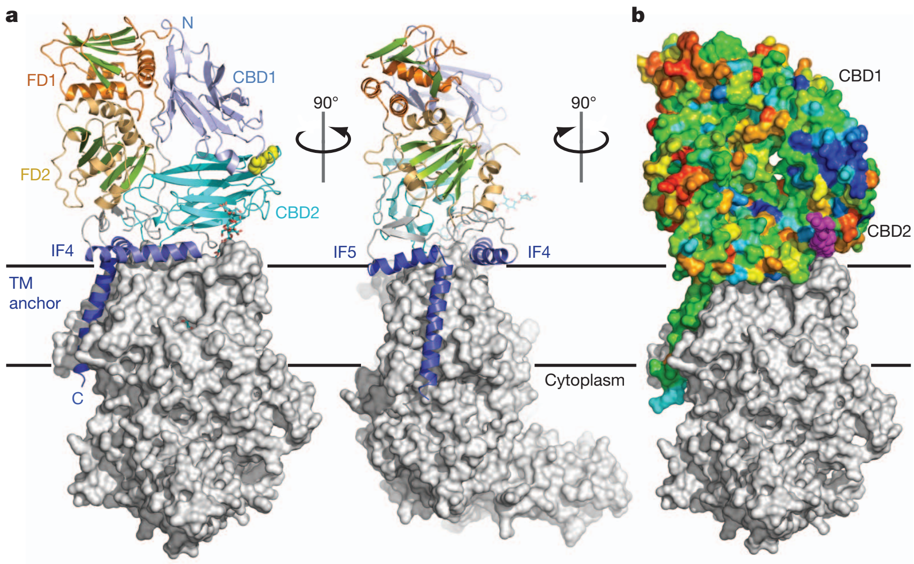


Figure 4 | Organization of the periplasmic domain. **a**, The periplasmic region of the cellulose synthase is primarily formed by BcsB. It consists of two periplasmic carbohydrate binding domains (CBD1 and CBD2) that are connected to two α/β domains (FD1 and FD2). The CBDs are covalently attached via a disulphide bond between the conserved Cys 163 and Cys 430 (yellow spheres). BcsB's transmembrane anchor and interface helices 4 and 5 are coloured in dark blue. The N and C termini of BcsB are indicated. **b**, Surface representation of BcsB coloured according to sequence conservation from red (variable) to deep blue (invariant). The tips of CBD1 and CBD2 form a patch of conserved, primarily hydrophobic residues above the periplasmic exit of the transmembrane channel. BcsA is shown as a grey surface and the glucan as purple spheres.

TM2 and TM3, as well as with its 5/6 loop, thereby accounting for almost 30% of the complex interface. Although BcsB exhibits significant sequence variability, we noticed a cluster of conserved residues including His 159, Arg 160, Ile 161, Leu 171 and Trp 172 at the tip of CBD1, close to its disulphide bond with CBD2 and above the periplasmic exit of the glucan channel (Fig. 4b). A comparison of CBD1 with the bacterial carbohydrate binding module family 35 (CBM35) in complex with a glucuronic acid disaccharide[31] localizes disaccharide binding to the cluster of conserved residues in CBD1 (Supplementary Fig. 9), indicating that this region interacts with the translocating glucan.

Both CBDs are linked to compact α/β domains containing a central β-sheet five or six strands wide that is framed on either side by two α-helices (Fig. 4a). Although probably lacking functional similarity, both domains resemble a flavodoxin fold[32] (Supplementary Fig. 10) and are thus referred to as FD1 (amino acids 192–303) and FD2 (amino acids 457–529 together with 598–666). The repeating structural motif of BcsB therefore contains a CBD linked to an FD domain.

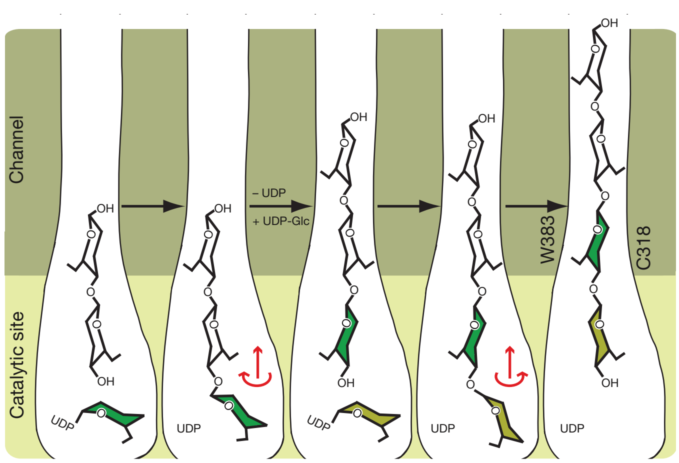


Figure 5 | Proposed model for cellulose synthesis and translocation. After glycosyl transfer, the newly added Glc could rotate around the acetal linkage into the plane of the polymer. The rotation direction would be determined by steric interactions and formation of the β-1,4 glucan characteristic intramolecular O3-H••O5 hydrogen bond. The glucan might translocate into the channel during this relaxation. This process would be repeated with a second UDP-Glc but the rotation direction after glycosyl transfer would be in the opposite direction owing to steric constraints. Alternatively, the glucan might not translocate into the channel until UDP is replaced by UDP-Glc. Trp 383 and Cys 318 mark the entrance to the transmembrane channel (only shown in the right panel).

Implications for cellulose synthesis and translocation

The structure of the cellulose synthase–cellulose translocation intermediate provides unique insights into the mechanism of cellulose synthesis and translocation. On the basis of the large distance between the glycosyltransferase site and the putative water–membrane interface (approximately 25 Å; Fig. 1) as well as BcsA's activity in detergent, a reaction mechanism involving lipid-linked intermediates seems unlikely[5]. The spatial restrictions in the substrate-binding pocket further suggest that the nascent glucan is extended by one[33] and not by two[34] glucose molecules at a time.

The distinct features of the glucan electron density (Fig. 1b) further indicate that the polymer translocates one glucose molecule at a time. The dimensions of the transmembrane channel require that the glucopyranose rings have to be within the same plane to slide into the pore (Fig. 3b). Thus, we speculate that after glycosyl transfer the newly attached terminal glucose molecule will rotate around the acetal linkage[35,36] to align with the glucan in the channel (Fig. 5). Steric interactions might dictate the rotation direction, leading to the β-1,4 glucan characteristic 180° rotation between individual glucose units and intramolecular hydrogen bonding between ring oxygens and 3′ hydroxyl groups of neighbouring units[4,37]. This relaxation might be sufficient to allow the polymer to slide into the channel (Fig. 5). Alternatively, the elongated glucan might not translocate until UDP is replaced with UDP-Glc, necessitating translocation owing to spatial restrictions at the active site. Past the channel, the induced kink of the glucan chain at the BcsA–BcsB interface, the interaction with BcsB's CBDs, or the aggregation with other glucans might further contribute to a unidirectional movement of the polymer.

Cellulose, chitin and hyaluronan are probably synthesized by a conserved mechanism involving a membrane-embedded glycosyltransferase that couples polymer synthesis with its translocation across the cell membrane[3]. This unique mechanism stands in contrast to the translocation of most other biological polymers where 'polymer synthases' function independently of or alongside dedicated translocation machineries[38–41]. Our structure now provides a basis for unravelling the details of this process.

METHODS SUMMARY

The BcsA and BcsB subunits of the *Rhodobacter sphaeroides* cellulose synthase complex were co-expressed and purified by metal affinity and gel filtration chromatography as described in detail in the Methods. The BcsA–BcsB complex was crystallized and the structure was determined after single anomalous dispersion phasing of a data set obtained from samarium III-chloride-soaked crystal. Additional experimental

phases were obtained from selenomethionine-derivatized BcsA–BcsB crystals. The quality of the electron density differed between peripheral and transmembrane regions of the complex. Therefore, modelling of the N-terminal region of BcsB (amino acids 54–177) and amino acids 586–675 of BcsA relied on the position of bulky side chains, Se-Met positions and backbone electron density. The model was built in Coot[42] and includes all residues with the exception of the terminal residues 1–12 and 760–788 of BcsA and 19–53 and 721–725 of BcsB. In addition, a short loop of BcsB consisting of residues 532–543 was not visible in the electron density map.

Full Methods and any associated references are available in the online version of the paper.

Received 12 July; accepted 31 October 2012.
Published online 9 December 2012.

1. Somerville, C. Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* **22,** 53–78 (2006).
2. Merzendorfer, H. Insect chitin synthases: a review. *J. Comp. Physiol. B* **176,** 1–15 (2006).
3. Hubbard, C., McNamara, J. T., Azumaya, C., Patel, M. S. & Zimmer, J. The hyaluronan synthase catalyzes the synthesis and membrane translocation of hyaluronan. *J. Mol. Biol.* **418,** 21–31 (2012).
4. Nishiyama, Y., Langan, P. & Chanzy, H. Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. *J. Am. Chem. Soc.* **124,** 9074–9082 (2002).
5. Matthysse, A. G., Thomas, D. L. & White, A. R. Mechanism of cellulose synthesis in Agrobacterium tumefaciens. *J. Bacteriol.* **177,** 1076–1081 (1995).
6. Bureau, T. E. & Brown, R. M. *In vitro* synthesis of cellulose II from a cytoplasmic membrane fraction of Acetobacter xylinum. *Proc. Natl Acad. Sci. USA* **84,** 6985–6989 (1987).
7. Grimson, M. J., Haigler, C. H. & Blanton, R. L. Cellulose microfibrils, cell motility, and plasma membrane protein organization change in parallel during culmination in Dictyostelium discoideum. *J. Cell Sci.* **109,** 3079–3087 (1996).
8. Kimura, S. & Itoh, T. New cellulose synthesizing complexes (terminal complexes) involved in animal cellulose biosynthesis in the tunicate *Metandrocarpa uedai*. *Protoplasma* **194,** 151–163 (1996).
9. Aloni, Y., Delmer, D. P. & Benziman, M. Achievement of high rates of *in vitro* synthesis of 1,4-beta-D-glucan: activation by cooperative interaction of the *Acetobacter xylinum* enzyme system with GTP, polyethylene glycol, and a protein factor. *Proc. Natl Acad. Sci. USA* **79,** 6448–6452 (1982).
10. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).
11. Römling, U. Molecular biology of cellulose production in bacteria. *Res. Microbiol.* **153,** 205–212 (2002).
12. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37,** D233–D238 (2009).
13. Standal, R. *et al.* A new gene required for cellulose production and a gene encoding cellulolytic activity in *Acetobacter xylinum* are colocalized with the bcs operon. *J. Bacteriol.* **176,** 665–672 (1994).
14. Saxena, I. M., Kudlicka, K., Okuda, K. & Brown, R. M. Characterization of genes in the cellulose-synthesizing operon (acs operon) of *Acetobacter xylinum*: implications for cellulose crystallization. *J. Bacteriol.* **176,** 5735–5752 (1994).
15. Hu, S. Q. *et al.* Structure of bacterial cellulose synthase subunit D octamer with four inner passageways. *Proc. Natl Acad. Sci. USA* **107,** 17957–17961 (2010).
16. Jahn, C. E., Selimi, D. A., Barak, J. D. & Charkowski, A. O. The *Dickeya dadantii* biofilm matrix consists of cellulose nanofibres, and is an emergent property dependent upon the type III secretion system and the cellulose synthesis operon. *Microbiology* **157,** 2733–2744 (2011).
17. Stewart, P. S. & Costerton, J. W. Antibiotic resistance of bacteria in biofilms. *Lancet* **358,** 135–138 (2001).
18. Amikam, D. & Galperin, M. Y. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* **22,** 3–6 (2006).
19. Römling, U., Gomelsky, M. & Galperin, M. Y. C-di-GMP: the dawning of a novel bacterial signalling system. *Mol. Microbiol.* **57,** 629–639 (2005).
20. Ross, P. *et al.* Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325,** 279–281 (1987).
21. Koyama, M., Helbert, W., Imai, T., Sugiyama, J. & Henrissat, B. Parallel-up structure evidences the molecular directionality during biosynthesis of bacterial cellulose. *Proc. Natl Acad. Sci. USA* **94,** 9091–9095 (1997).
22. Lai-Kee-Him, J. *et al. In vitro* versus *in vivo* cellulose microfibrils from plant primary wall synthases: structural differences. *J. Biol. Chem.* **277,** 36931–36939 (2002).
23. Charnock, S. J. & Davies, G. J. Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38,** 6380–6385 (1999).
24. Weigel, P. H. & Deangelis, P. L. Hyaluronan synthases: a decade-plus of novel glycosyltransferases. *J. Biol. Chem.* **282,** 36777–36781 (2007).
25. Kozmon, S. & Tvaroska, I. Catalytic mechanism of glycosyltransferases: hybrid quantum mechanical/molecular mechanical study of the inverting N-acetylglucosaminyltransferase I. *J. Am. Chem. Soc.* **128,** 16921–16927 (2006).
26. Kubota, T. *et al.* Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide α-*N*-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J. Mol. Biol.* **359,** 708–727 (2006).
27. Daras, G. *et al.* The thanatos mutation in *Arabidopsis thaliana* cellulose synthase 3 (AtCesA3) has a dominant-negative effect on cellulose synthesis and plant growth. *New Phytol.* **184,** 114–126 (2009).
28. Benach, J. *et al.* The structural basis of cyclic diguanylate signal transduction by PilZ domains. *EMBO J.* **26,** 5153–5166 (2007).
29. Scheible, W. R., Eshed, R., Richmond, T., Delmer, D. & Somerville, C. Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis* Ixr1 mutants. *Proc. Natl Acad. Sci. USA* **98,** 10079–10084 (2001).
30. Harris, D. M. *et al.* Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1A903V and CESA3T942I of cellulose synthase. *Proc. Natl Acad. Sci. USA* **109,** 4098–4103 (2012).
31. Montanier, C. *et al.* Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proc. Natl Acad. Sci. USA* **106,** 3065–3070 (2009).
32. Sancho, J. Flavodoxins: sequence, folding, binding, function and beyond. *Cell. Mol. Life Sci.* **63,** 855–864 (2006).
33. Delmer, D. P. Cellulose biosynthesis: Exciting times for a difficult field of study. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50,** 245–276 (1999).
34. Carpita, N. C. Update on mechanisms of plant cell wall biosynthesis: How plants make cellulose and other (1→4)-β-D-glycans. *Plant Physiol.* **155,** 171–184 (2011).
35. Dowd, M. K., French, A. D. & Reilly, P. J. Conformational analysis of the anomeric forms of sophorose, laminarabiose, and cellobiose using MM3. *Carbohydr. Res.* **233,** 15–34 (1992).
36. Momany, F. A. & Schnupf, U. DFTMD studies of β-cellobiose: conformational preference using implicit solvent. *Carbohydr. Res.* **346,** 619–630 (2011).
37. Brown, M. Cellulose structure and biosynthesis: What is in store for the 21st century? *J. Pol. Sci.* **42,** 487–495 (2004).
38. Rapoport, T. A. Protein transport across the endoplasmic reticulum membrane. *FEBS J.* **275,** 4471–4478 (2008).
39. Burton, B. & Dubnau, D. Membrane-associated DNA transport machines. *Cold Spring Harb. Perspect. Biol.* **2,** 1–20 (2010).
40. Whitfield, C. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* **75,** 39–68 (2006).
41. Raetz, C. R. H. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71,** 635–700 (2002).
42. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to G. Murshudov for advice on Refmac refinement and U. and Z. Derewenda for discussions. We thank L. Tamm, M. Wiener, A. Walling and T. Rapoport for critical comments on the manuscript. X-ray diffraction data were collected at GM/CA- and Southeast Regional-Collaborative Access Team beamlines at the Advanced Photon Source (APS), Argonne National Laboratory. Use of the APS was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, contract no. DE-AC02-06CH11357 and W-31-109-Eng-38. GM/CA at APS has been funded in whole or in part with funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). The University of Georgia CCRC is supported by the Department of Energy funded Center for Plant and Microbial Complex Carbohydrates (DE-FG02-09ER-20097). J.L.W.M. is partially supported by a Peach Fellowship, University of Virginia. J.Z. is support by NIH grant 1R01GM101001 and start-up funds from the University of Virginia School of Medicine.

Author Contributions J.Z. designed the experiments. J.L.W.M. and J.Z. expressed, purified and crystallized the BcsA–BcsB complex. J.L.W.M. and J.Z. analysed the crystallographic data and built the model. J.S. performed *in vitro* cellulose synthesis assays. J.L.W.M. and J.Z. wrote the manuscript.

Author Information The atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession number 4HG6. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.Z. (jochen_zimmer@virginia.edu).

METHODS

Expression and purification of BcsA–BcsB. The *bcsA* gene was PCR amplified from *Rhodobacter sphaeroides* 2.4.1. genomic DNA (ATCC) and cloned using NcoI and HindIII restriction sites into an engineered pETDuet expression vector (Novagen) conferring a C-terminal dodeca-histidine tag. The *bcsB* gene was PCR amplified without its native, N-terminal signal sequence. The gene coding for amino acids 21–725 of BcsB was cloned into the pETDuet vector containing the *bcsA* gene, together with an N-terminal PelB signal sequence using NdeI and KpnI restriction sites. After signal sequence cleavage, this generates an additional Met-Gly dipeptide at the N terminus of BcsB. Electro-competent Rosetta 2 cells were transformed with the pETDuetBcsA-12His-BcsB construct and plated on agar plates supplemented with 100 μg ml^{-1} ampicillin and 25 μg ml^{-1} chloramphenicol and grown overnight at 37 °C. The transformed cells were re-suspended and used to inoculate 36 l of ZYP-5052 auto-induction medium[43]. The cells were grown shaking at 37 °C for at least 4 h after the optical density at 600 nm reached 0.8. Subsequently, the cells were collected by centrifugation, re-suspended in RB1-buffer containing 20 mM sodium phosphate buffer pH 7.2, 0.1 M NaCl and 5 mM cellobiose and then lysed in a microfluidizer. The crude membranes were collected by centrifugation for 60 min at 120,000*g* in a Beckman Ti45 rotor and solubilized for 60 min at 4 °C in RB2 buffer containing 20 mM sodium phosphate buffer pH 7.2, 0.3 M NaCl, 5 mM cellobiose, 5 mM $MgCl_2$, 40 mM imidazole, 10% glycerol and 2% Triton X-100. The insoluble material was cleared by centrifugation for 30 min at 120,000*g* in a Beckman Ti45 rotor and the membrane extract was batch incubated with 10 ml Ni-NTA agarose (Qiagen) for 45 min at 4 °C. The resin was packed in a gravity flow chromatography column, washed with 75 ml WB1-buffer (RB2-buffer containing a total of 60 mM imidazole and 5 mM lauryldimethylamine-*N*-oxide, LDAO), 25 ml WB2-buffer (RB2 buffer containing a total of 1 M NaCl and 5 mM LDAO) and the BcsA–BcsB complex was eluted in 50 ml EB-buffer containing 20 mM Tris HCl pH 7.5, 0.1 M NaCl, 5 mM cellobiose, 5 mM $MgCl_2$, 250 mM imidazole, 10% glycerol and 5 mM LDAO. The eluted protein was concentrated and purified over an analytical S200 gel filtration column (GE Healthcare) equilibrated in EB buffer containing 20 mM MES pH 6.5 instead of Tris HCl and lacking $MgCl_2$ and imidazole (GF-buffer). The eluted BcsA–BcsB was concentrated to 50 μM final concentration (using an extinction coefficient at 280 nm of 161,925 $M^{-1} cm^{-1}$) in a 100-kDa cut-off centrifugal filter (Millipore) and supplemented with 10 mM EDTA (EDTA) and 1 mM uridine diphosphate-glucose (UDP-Glc) before crystallization.

Selenomethionine-derivatized BcsA–BcsB was prepared as described above with the exception that the cells were grown in the auto-inducing PASM-5052 minimal medium supplemented with 125 μg ml^{-1} selenomethionine (Se-Met)[43].

The BcsA–BcsB complex was crystallized in 30% PEG200, 0.1 M MES pH 6.5 and 50 mM NaCl at 4 °C. Initial crystals were observed after approximately 7 days and reached their final size of about 50 × 50 × 100 μm within 3 weeks. The crystals were collected and directly cryo-cooled in liquid nitrogen. Wild-type crystals were soaked with 20 mM samarium III-chloride (Sm) or 1 mM sodium ethylmercuric-thiosalicylate (EMTS) for 2–24 h before collection.

Structure determination. X-ray diffraction data were collected at the Argonne National Laboratory beamlines SER- as well as GM/CA-CAT. The data were indexed and integrated in Mosflm[44], analysed in Pointless and scaled and reduced in space group $P4_32_12$ in Scala from the CCP4 suite of programs[45]. The Sm-derivatized crystals diffracted to 5 Å resolution and were used for single-wavelength anomalous dispersion (SAD) phasing in ShelX[46]. The initial phases allowed localization of all transmembrane helices of BcsA and were further used to calculate an anomalous difference Fourier electron density map for the data set obtained from Se-Met-derivatized BcsA–BcsB. Initially, ten Se-Met positions were identified and refined in Mlphare[45]. The phases were combined with the native data set followed by density modification with phase extension to 3.25 Å in DM[47] using a user-generated solvent mask. Throughout the model building process, Se-Met anomalous difference Fourier electron density maps were calculated, additional Se-Met positions were refined as described above and new phases were calculated and improved by density modification in Parrot[47]. The final experimental phases were calculated from 36 Se-Met-, 4 EMTS- and 8 Sm positions and allowed tracing of most of the backbone of BcsA and the membrane proximal regions of BcsB. The model was built in Coot[42] via an iterative process involving careful model building, combination of experimental and model phases, density modification and restraint refinement in Refmac-5[45] (including TLS parameters[48] for the final model). The model includes a β-1,4-linked glucan containing 18 glucose molecules and a UDP molecule, as well as 2 partially ordered LDAO molecules. The UDP molecule was refined with 50% occupancy owing to weak electron density. The refined model has 91.8% of its residues in the favoured region of the Ramachandran diagram and includes 1.1% outliers, as determined by MolProbity. All figures were prepared in Pymol[49] and the sequence conservation score for BcsB was calculated using CONSURF[50].

***In vitro* cellulose synthesis.** 100 μl of BcsA–BcsB in GF buffer was incubated with 5 mM UDP-Glc, 20 mM $MgCl_2$ and 30 μM cyclic-di-GMP at 37 °C for 90 min. The polymerization reaction was terminated with 40 mM EDTA and the sample was centrifuged at 14,000*g* at room temperature for 20 min. The supernatant was discarded and the pellet was washed twice in 500 μl deionized water, dried and subjected to linkage analysis by the Complex Carbohydrate Research Center, University of Georgia.

43. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
44. Leslie, A. G. W. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62,** 48–57 (2006).
45. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
46. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64,** 112–122 (2008).
47. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D* **66,** 470–478 (2010).
48. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62,** 439–450 (2006).
49. PyMol. The PYMOL Molecular Graphics System (DeLano Scientific) (2000).
50. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** W529–W533 (2010).

DEVELOPMENTAL BIOLOGY

Led by the nose

Hagfish embryos show developmental features that contradict the idea that these jawless fish are the most primitive living vertebrates. The findings also help to trace the evolution of vertebrate cranial structure. SEE ARTICLE P.175

PHILIPPE JANVIER

Hagfish are bizarre creatures. These eel-shaped marine fish have a skull but no vertebrae, and teeth but no jaw. These and other odd features have long fascinated evolutionary biologists, who regard the 'primitive' characteristics of hagfish as a treasure trove of clues to understanding the early evolution of vertebrate anatomy and physiology. But hagfish embryos are reputedly scarce, so their development has remained an enigma. Six years ago, however, a team of Japanese researchers realized[1] that hagfish embryos are not particularly rare: they simply take more time to mature than other piscine vertebrates. This discovery triggered extensive investigations of hagfish embryos, some of which are published by Oisi *et al.*[2] on page 175 of this issue*. The authors provide a spectacular overview of hagfish craniofacial development, with particular reference to the origins of the olfactory organs and pituitary gland, two key features in the shaping of the vertebrate skull[3].

Hagfish often feed on sunken carcasses, tearing off food using a peculiar tongue-like feeding device that is armed with sharp teeth made of keratin. They were first classified as cyclostomes, alongside lampreys, because both groups of jawless fish have similar feeding devices and pouch-shaped gills, and both lack a calcified skeleton. Cyclostomes were regarded as an early offshoot of the vertebrate evolutionary tree. Then, in the 1970s, hagfish were considered to be anatomically and physiologically more primitive than either lampreys or jawed vertebrates (gnathostomes), and to have diverged earlier than all other vertebrates. However, this hypothesis has since been dismissed on the basis of DNA sequence data that show a close phylogenetic relationship between lampreys and hagfish[4,5]. But until recently, lamprey development had been studied much more than that of hagfish.

Hagfish and lampreys have a single median (located in the midline of the body) 'nostril', which opens anteriorly in hagfish but dorsally in lampreys. This nostril is the entrance of a tube into which two organs open: a median olfactory organ and the adenohypophysis (also known as the anterior pituitary gland, which is part of the endocrine system). This monorhinal (single nostril) condition differs from the diplorhinal (two nostrils) structure of the gnathostomes, in which two well-separated olfactory organs open to the exterior environment through independent nostrils, as do human noses. The separation of the olfactory organs in gnathostomes is now regarded as a prerequisite to the development of jaws[3], because it allowed the mandibular arch to extend forward.

However, this difference in the structures poses a conundrum. The cyclostomes and gnathostomes are each monophyletic[4,5], meaning that they form a group comprising an ancestral species and all its descendants. Non-vertebrate members of the phylum Chordata (to which vertebrates also belong), such as amphioxus or tunicates (sea squirts), lack an indisputable homologue of the olfactory organs and adenohypophysis. This raises the question: what was the arrangement of these organs in the most recent common ancestor to all vertebrates? To answer this, Oisi *et al.* studied the embryonic tissues that give rise to the olfactory organs and adenohypophysis in hagfish, lampreys and gnathostomes. These tissues, called placodes, are thickenings of the embryonic ectoderm (the outer germ layer) that develop into various sensory organs, such as the olfactory organs, eye lens, otic capsules and the adenohypophysis.

In the lamprey embryo, the olfactory and adenohypophyseal placodes are united into a single nasohypophyseal plate (NHP) that forms beneath the upper lip and gives rise to a nasohypophyseal complex, which migrates dorsally later in development (Fig. 1). Oisi and colleagues found that hagfish embryos also have an NHP, and that the gene-expression pattern in both the lamprey and hagfish NHP is characteristic of ectoderm-associated tissues. These findings clearly refute previous speculation that the hagfish NHP derives from the endoderm (the inner germ layer that later forms the pharynx and digestive tract), which, if true, would have been the only case of this among vertebrates.

But this reconciliation of the structure and position of the NHP in hagfish and lampreys still leaves two distinct conditions among vertebrates: cyclostomes have a median NHP, and gnathostomes have independent nasal and adenohypophyseal placodes (Fig. 1).

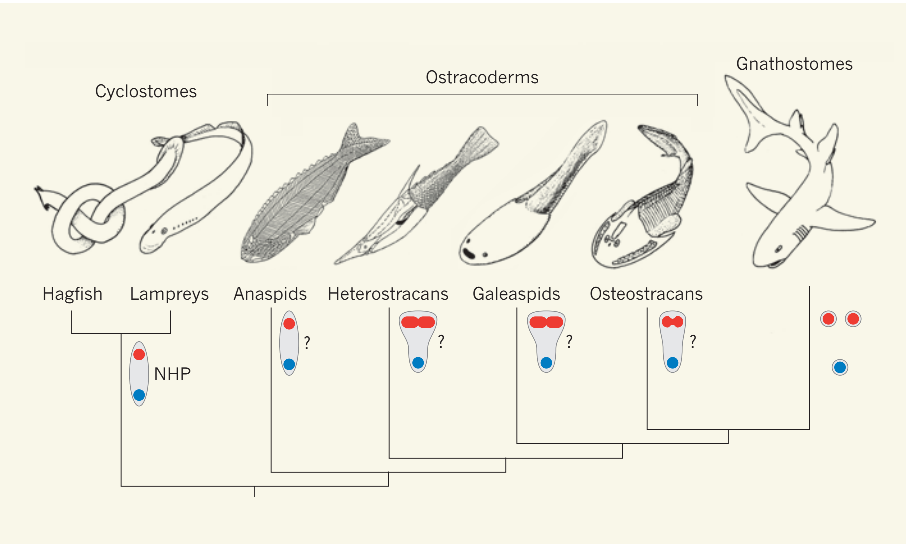


Figure 1 | Cranial anatomy of living and fossil vertebrates. Oisi *et al.*[2] show that the olfactory organs and anterior pituitary gland (adenohypophysis) of the two extant forms of jawless vertebrates, hagfish and lampreys (collectively referred to as the cyclostomes), develop in much the same way. Their embryos have median olfactory (red) and adenohypophyseal (blue) placodes (thickenings of the outer germ layer) that are united in a nasohypophyseal plate (NHP). By contrast, in jawed vertebrates (gnathostomes), these organs arise from separate tissues: a pair of olfactory placodes and a single adenohypophyseal placode. Although placodes are embryonic structures not seen in fossils, the skeletons of extinct jawless fish (ostracoderms), which are regarded as more closely related to gnathostomes than to cyclostomes, can help in reconstructing the steps towards gnathostome anatomy (the drawings are reconstructions from fossil evidence[6]). The anaspids and osteostracans, for example, display a small dorsal nostril similar to that of lampreys, whereas the heterostracans and galeaspids had paired olfactory organs and a large shared nostril. This diversity foreshadows the transition from single-nostrilled to paired-nostrilled forms, and the separation of olfactory and adenohypophyseal development during vertebrate evolution may have allowed the forward extension of the mandibular arch and hence the development of jaws.

*This article and the paper under discussion[2] were published online on 19 December 2012.

Could one of these conditions be general — or 'ancestral' — for vertebrates, and the other have subsequently derived from it? Oisi *et al.* suggest that the cyclostome condition (a median NHP) is likely to be the earlier structure, because the monorhinal condition is found in the extinct jawless vertebrates commonly known as ostracoderms, which are between 480 million and 370 million years old and are regarded as more closely related to gnathostomes than to cyclostomes (Fig. 1).

Placodes are embryonic structures that are not preserved in ostracoderm fossils, but the skeletons of these animals provide information about their differing cranial anatomy[6,7]. Ostracoderms display diverse arrangements of the cranial spaces that housed the olfactory organs and the adenohypophysis, but in all of them these spaces open towards a median nostril, as in cyclostomes. For example, anaspids and osteostracans have a median dorsal nasohypophyseal complex similar to that of lampreys, which suggests some kind of cyclostome-like NHP, although other features show that osteostracans are closer to gnathostomes than to any other ostracoderm. By contrast, heterostracans and galeaspids had paired olfactory organs, still closely associated with the adenohypophysis in the latter, but opening anteriorly by a large median nostril and suggesting a broader NHP (Fig. 1).

This anatomical diversity is puzzling, but Oisi and colleagues' findings on hagfish development help to clarify the picture. The authors propose that the presence of the NHP in both hagfish and lamprey embryos supports the idea of a general, 'pan-cyclostome' pattern for craniofacial development, and that the anatomical characteristics that make present-day hagfish seem more primitive than lampreys are the result of a subsequent loss of features in hagfish during evolution. Such a pattern, combined with the fact that no NHP is seen during the development of any living gnathostomes, and the apparent progression from the monorhinal to diplorhinal condition in ostracoderms, suggests that a crucial innovation in the development of the vertebrate snout was the loss of a common nasohypophyseal duct and the rise of separate, paired nostrils. ■

Philippe Janvier *is at the Muséum National d'Histoire Naturelle, 75231 Paris Cedex 05, France.*
e-mail: janvier@mnhn.fr

1. Ota, K. G. & Kuratani, S. *Zool. Sci.* **23,** 403–418 (2006).
2. Oisi, Y., Ota, K. G., Kuraku, S., Fujimoto, S. & Kuratani, S. *Nature* **493,** 175–180 (2013).
3. Kuratani, S. *J. Anat.* **205,** 335–347 (2004).
4. Kuraku, S. *Zool. Sci.* **25,** 960–968 (2008).
5. Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. & Peterson, K. J. *Proc. Natl Acad. Sci. USA* **105,** 2946–2950 (2008).
6. Janvier, P. *Zool. Sci.* **25,** 1045–1056 (2008).
7. Gai, Z., Donoghue, P. C. J., Zhu, M., Janvier, P. & Stampanoni, M. *Nature* **476,** 324–327 (2011).

PHOTONICS

Phased array on a fingertip

An array of more than 4,000 optical antennas working in unison has been demonstrated on a millimetre-scale silicon chip. The result highlights the remarkable capabilities of optical integration in silicon. SEE LETTER P.195

THOMAS F. KRAUSS

In his Nobel lecture in 1909, physicist Ferdinand Braun remarked[1] that: "It had always seemed most desirable to me to transmit the waves, in the main, in one direction only". He was referring to radio waves, which had hitherto been known to radiate only in a doughnut shape away from the emitter, as described by Heinrich Hertz two decades earlier. The invention for which Braun was recognized was the phased array — the idea of connecting up multiple antennas fed from a common source and using the relationship between the antennas' phase (where a wave's peaks and troughs lie) to enhance the emission of radio waves in a given direction. His work was instrumental in the development of radio transmitters, and led to other inventions such as the television antenna and radar. The phased-array concept is now also extensively used in astronomy, where the information from multiple telescopes is collected in phase to extend the resolution of the system. On page 195 of this issue, Sun *et al.*[2] introduce an integrated optical phased array that combines the equivalent of more than 4,000 telescopes onto a single chip smaller than a fingernail.

The creation of phased arrays in the optical domain has been an active area of research for some time, as it is motivated by the need to improve applications such as optical beam steering and imaging. To achieve the highest performance, the challenges are to integrate a large number of antennas onto a single chip, to place them close together and to fully control the phase on each element. The short optical wavelength makes it easy, in principle, to address the first challenge and to build large arrays on a small footprint in an integrated format. The other two challenges are much more difficult to meet, however, because the optical couplers that feed a controlled amount of power into each antenna and the phase shifters needed to control the phase on each element tend to be tens or hundreds of wavelengths long, which requires a large distance between each antenna. Therefore, the latest realizations of integrated optical phased arrays have been relatively large and have featured a maximum of only 4×4 antennas[3,4].

The architecture now introduced by Sun *et al.*[2] addresses all of these requirements and is scalable. By carefully designing the optical coupler, phase shifter and gratings that act as radiating antennas, the team has managed to produce a phased array of 64×64 antenna units, or 'pixels', on a footprint of a little more than 0.576×0.576 millimetres, with each pixel covering 9×9 micrometres — equivalent to a pixel side length of six optical wavelengths. Most impressive are the accuracy and control that the researchers were able to exercise over the array's optical functions, especially those of the optical couplers.

The device's accuracy and performance are highlighted by the well-balanced power output achieved across the array. Equally, the fact that a phase shift of π radians can be attained on a pixel size a few wavelengths across (in contrast to the tens or hundreds of wavelengths previously required), and with relatively low power and crosstalk, points to the remarkable capabilities of nanophotonic concepts that are now available. Fabrication of the array, as a photonic circuit, using only silicon microelectronics processing techniques — a concept pioneered[5] by the IMEC microelectronics centre in Leuven, Belgium — is also a testimony to the high performance that can be achieved by using silicon as a photonic material. The successful demonstration of the authors' array shows that the large-scale integration of photonic circuitry in silicon has well and truly arrived.

So where do we go from here? One interesting idea is that of imaging through light-scattering media[6], such as roughened glass or biological tissue. Such imaging uses the adaptive-optics principle of adjusting the phase of an optical wave locally so as to compensate for the distortion caused by the medium. The authors' optical phased array now makes available the phase control and large pixel number that this application requires. To achieve imaging through time-varying light scatterers, such as turbulent liquids, a fast time-response of the system is also essential; the small and efficient tuning of each antenna demonstrated by Sun *et al.* can also meet this requirement. Another application is in sensor networks, in which communication between randomly distributed sensors is based on

LETTER

doi:10.1038/nature11769

Flows of gas through a protoplanetary gap

Simon Casassus[1], Gerrit van der Plas[1], Sebastian Perez M[1], William R. F. Dent[2,3], Ed Fomalont[4], Janis Hagelberg[5], Antonio Hales[2,4], Andrés Jordán[6], Dimitri Mawet[3], Francois Ménard[7,8], Al Wootten[4], David Wilner[9], A. Meredith Hughes[10], Matthias R. Schreiber[11], Julien H. Girard[3], Barbara Ercolano[12], Hector Canovas[11], Pablo E. Román[13] & Vachail Salinas[1]

The formation of gaseous giant planets is thought to occur in the first few million years after stellar birth. Models[1] predict that the process produces a deep gap in the dust component (shallower in the gas[2–4]). Infrared observations of the disk around the young star HD 142527 (at a distance of about 140 parsecs from Earth) found an inner disk about 10 astronomical units (AU) in radius[5] (1 AU is the Earth–Sun distance), surrounded by a particularly large gap[6] and a disrupted[7] outer disk beyond 140 AU. This disruption is indicative of a perturbing planetary-mass body at about 90 AU. Radio observations[8,9] indicate that the bulk mass is molecular and lies in the outer disk, whose continuum emission has a horseshoe morphology[8]. The high stellar accretion rate[10] would deplete the inner disk[11] in less than one year, and to sustain the observed accretion matter must therefore flow from the outer disk and cross the gap. In dynamical models, the putative protoplanets channel outer-disk material into gap-crossing bridges that feed stellar accretion through the inner disk[12]. Here we report observations of diffuse CO gas inside the gap, with denser HCO$^+$ gas along gap-crossing filaments. The estimated flow rate of the gas is in the range of 7×10^{-9} to 2×10^{-7} solar masses per year, which is sufficient to maintain accretion onto the star at the present rate.

The HD 142527 system offers an unhindered view of its large central gap, and is a promising 'laboratory' in which to observe the ongoing formation of gaseous giant planets. The orientation of the disk is well understood. Multiwavelength data are consistent with an inclination of about 20°, indicating that the disk is almost face-on to our line of sight[11]. The disk position angle is about −20° east of north, and the eastern side is the far side of the disk, as suggested by a clear view of the outer disk's inner rim in the mid-infrared spectral range[11,13] and by a clockwise rotation suggested by what is probably a trailing spiral arm to the west[6].

We find that the CO(3–2) emission (that is, emission of CO with $J = 3$–2, where J is the CO rotational quantum number) peaks inside the gap. Other disks have been observed to have a CO decrement within dust cavities[14,15], and may represent later evolutionary stages or different gap-clearing mechanisms. Gas inside dust cavities has previously been directly observed very close to the central star (inside the dust evaporation radius) using near-infrared interferometry[16–18]. Other indirect observations of gas inside[19–23] dust gaps at larger distances from the central star have revealed spectroscopically resolved gas tracers, such as rovibrational CO emission at a wavelength of 4.67 μm and [O I] emission at 6,300 Å, under the assumption of azimuthal symmetry and Keplerian rotation. Spectro-astrometry in combination with Keplerian disk models and azimuthal symmetry have also been used to infer the presence of CO gas inside disk gaps[24–26]. Our data provide a well-resolved observation of gas at sub-millimetre wavelengths inside a dust gap.

The dust gap that we see in the radio continuum (Fig. 1a) indicates a decrease by a factor of at least 300 in the surface density of millimetre-sized grains, judging from the contrast ratio between the peak on the horseshoe-shaped outer disk (the northern flux peak at 360 mJy per beam) and the faintest detected signal inside the gap (namely the western filament at 1 mJy per beam; see below). Yet there is no counterpart in the CO(3–2) map (Fig. 1b) of the arc that we see in the radio continuum. CO(3–2) is probably optically thick, as reflected by its diffuse morphology, so that it traces the temperature profile rather than the underlying density field. Detailed modelling of optically thin isotopologue data is required to constrain accurately the depth of the gaseous gap. To study the distribution of dense gas inside the gap, we use the tracer HCO$^+$.

The second result from our observations is that gas showing HCO$^+$(4–3) emission, expected in the denser regions (molecular hydrogen number density, $n_{H_2} \approx 10^6\,cm^{-3}$) exposed to ultraviolet radiation, is indeed found in the exposed rim of the dense outer disk, but also along gap-crossing filaments. The most conspicuous filament extends eastwards from the star, and a fainter filament extends westwards. Both filaments subtend an angle of about 140 ± 10° with the star at its vertex. The central regions of these filaments correspond to the brightest features in the HCO$^+$ line intensity maps (Fig. 1c), although the outer disk is brighter in peak specific intensity (Supplementary Fig. 7). Thus, line velocity profiles are broader on the stellar side of the filaments than in the outer disk, where they merge with the outer-disk Keplerian rotation pattern. These narrow, systemic-velocity HCO$^+$ filaments are best seen in intensity maps integrated over the filament velocities (Fig. 1d, inset). No central peak is seen in the channel maps (Supplementary Fig. 2), and so a beam-elongation effect can be ruled out. The eastern filament also is notable in terms of peak HCO$^+$ specific intensity (Supplementary Fig. 7e). For ease of visualization, we show deconvolved models of the HCO$^+$ intensity images in the inset to Fig. 1d. A related feature is seen in CO(3–2) emission, whose intensity peaks in the more diffuse regions surrounding the eastern HCO$^+$ filament. We note from the inset to Fig. 1 that the continuum also shows features under the HCO$^+$ filament. These features are faint and grow away from the edges of the horseshoe-shaped outer disk. Estimates of physical conditions are given in Supplementary Information.

The molecular and filamentary flows near the star are non-Keplerian. Blueshifted emission extends to the east from the central intensity peak (Fig. 1d). This velocity component is broad near the star, with emission ranging from −3.4 to +11 km s^{-1} (Supplementary Fig. 2), and is marginally resolved (the central HCO$^+$ peak extends over about 0.65×0.38 arcsec^2). In the deconvolved images of the inset to Fig. 1d, the peak intensity in the blue- and red-outlined regions are separated by about 0.2 arcsec, that is, by the diameter of the inner disk,

[1]Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile. [2]Joint ALMA Observatory, Alonso de Córdova 3107, Vitacura 763-0355, Santiago, Chile. [3]European Southern Observatory, Casilla 19001, Vitacura, Santiago, Chile. [4]National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, Virginia 22903-2475, USA. [5]Observatoire de Genève, Université de Genève, 51 Chemin des Maillettes, 1290, Versoix, Switzerland. [6]Departamento de Astronomía y Astrofísica, Pontificia Universidad Católica de Chile, Santiago, Chile. [7]UMI-FCA, CNRS/INSU France (UMI 3386), and Departamento de Astronomía, Universidad de Chile, Santiago, Chile. [8]CNRS/UJF Grenoble 1, UMR 5274, Institut de Planétologie et dAstrophysique de Grenoble (IPAG), F-48041 Grenoble Cedex 9, France. [9]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [10]Department of Astronomy, UC Berkeley, 601 Campbell Hall, Berkeley, California 94720, USA. [11]Departamento de Física y Astronomía, Universidad Valparaiso, Avenida Gran Bretana 1111, Valparaiso, Chile. [12]University Observatory, Ludwig-Maximillians University, D-81679 Munich, Germany. [13]Center of Mathematical Modeling, University of Chile, Avenida Blanco Encalada 2120 Piso 7, Santiago, Chile.

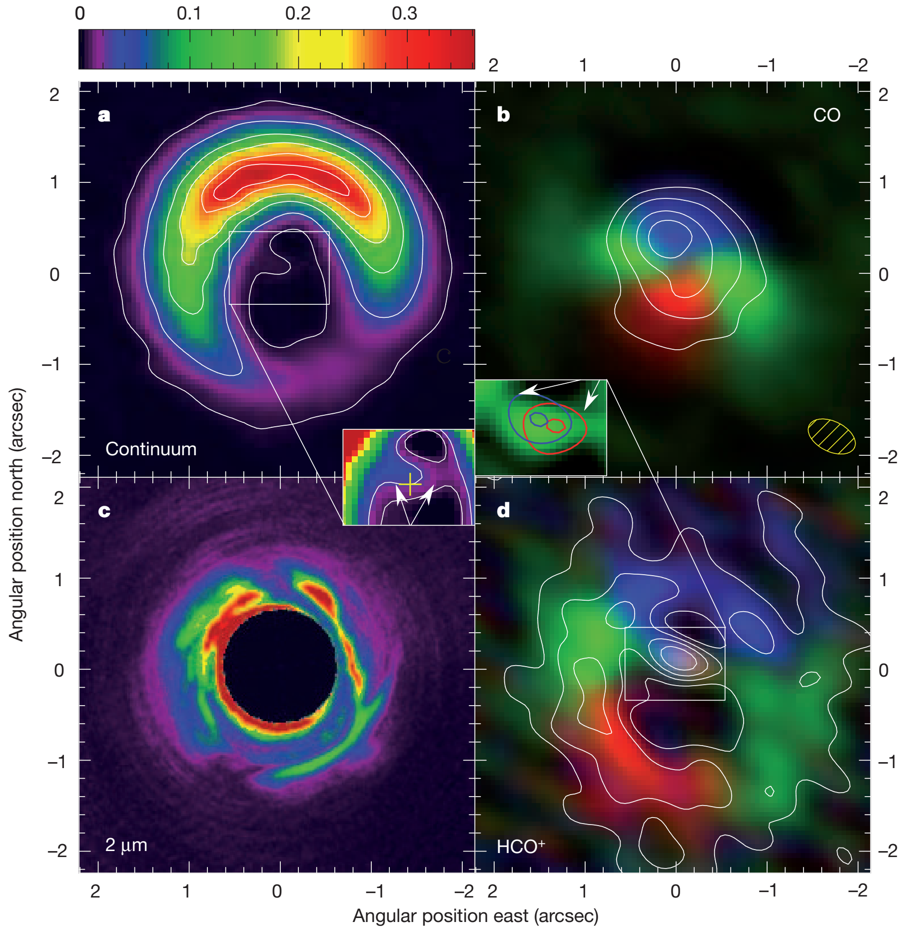


Figure 1 | ALMA observations of HD 142527, with a horseshoe dust continuum surrounding a gap that still contains gas. We see diffuse CO gas in Keplerian rotation (coded in Doppler-shifted colours), and filamentary emission in HCO^+, with non-Keplerian flows near the star (comparison models illustrative of Keplerian rotation are shown in Supplementary Information). The near-infrared emission abuts the inner rim of the horseshoe-shaped outer disk. The star is at the origin of the coordinates. North is up and east is to the left. **a**, Continuum at 345 GHz, with specific intensity units in Janskys per beam. It is shown on an exponential scale (colour scale). A beam ellipse is shown in the bottom right of **b**, and contours are drawn at 0.01, 0.1, 0.3, 0.5, 0.75 and 0.9 times the peak value. The noise level is $1\sigma = 0.5$ mJy per beam. **b**, CO(3–2) line intensity shown by white contours at fractions of 0.3, 0.5, 0.75 and 0.95 of the peak intensity value, 2.325×10^{-20} W per beam. The underlying red–green–blue image also shows CO(3–2) line intensity, but integrated in three different velocity bands, whose velocity limits are indicated in the spectra of Supplementary Fig. 8. **c**, Near-infrared image from Gemini that traces reflected stellar light, shown on a linear scale. We applied a circular intensity mask to the stellar glare, some of which immediately surrounds the mask. See Supplementary Fig. 1 for an overlay with the continuum. **d**, HCO^+(4–3) line intensity shown by white contours at fractions of 0.1, 0.3, 0.5, 0.75 and 0.95 of the peak intensity value, 0.40×10^{-20} W per beam, overlaid on a red–green–blue image of HCO^+ intensity summed in three different colour bands (see Supplementary Fig. 8 for definitions). Insets show magnified views of the central features that cross the dust gap. The cross indicates the star at the origin, with a precision of 0.05 arcsec, and the arrows point at the filaments. Inset to **a**: same as in **a**, with a narrow exponential scale highlighting the gap-crossing filaments. These features appear to grow from the eastern and western sides of the horseshoe. Contours are at 0.0015 and 0.005 Janskys per beam. Inset to **d**: deconvolved models (Supplementary Information) of the HCO^+ emission (green) at velocities where the gap-crossing filaments are seen, that is, from 3.2 to 4.3 km s^{-1}. Intensity maps for the blue and red velocity ranges (see Supplementary Fig. 8 for definitions) are shown in contours, with levels at 0.5 and 0.95 times the peak values. These red and blue contours are an alternative way to present the intensity field shown in **d**, but deconvolved for ease of visualization.

and at a position angle orthogonal to that expected from close-in high-velocity material in Keplerian rotation. Very blueshifted emission could reach out to 0.2 arcsec from the star (channel at -2.4 km s^{-1} in Supplementary Fig. 2, taking into account the beam). A blueshifted CO(3–2) high-velocity component can also be seen at the base of this feature, near the star (at -2.1 km s^{-1} in Supplementary Fig. 5).

The non-Keplerian HCO^+ is probably not consistent with a central outflow. Stellar outflows are not observed[27] in disks with inner cavities and no molecular envelopes (that is, transition disks). For an outflow orientation, the low velocities measured by the lines imply that the filaments in HD 142527 would stand still and hover above the star (Supplementary Information, section 3). Even the blueshifted emission is slow by comparison with the escape velocity. A slow disk wind (for example one photoevaporative or magnetically driven) can also be excluded on the basis of the high collimation shown by the HCO^+ emission. Indeed, the CO 4.67-µm emission seen in the inner disk[26] is purely Keplerian, it does not bear the signature of the disk winds seen in other systems and its orientation is the same as that of the outer disk. An orthogonal inner disk can also be discounted on dynamical grounds (Supplementary Information, section 3).

It is natural to interpret the filaments as planet-induced gap-crossing accretion flows, or 'bridges'. Because the eastern side is the far side, the blueshifted part of the eastern bridge is directed towards the star and is a high-velocity termination of the accretion flow onto the inner disk. These bridges are predicted by hydrodynamical simulations when applied to planet-formation feedback in HD 142527 (ref. 7). In these simulations, the bridges straddle the protoplanets responsible for the dynamical clearing of the large gap in HD 142527. They are close to Keplerian rotation in azimuth, but have radial velocity components of $\gtrsim 0.1$ of the azimuthal components. In our data, we see that

as the bridges contact the inner disk they also coincide with higher-velocity material, at two-dimensional radial velocities less than their azimuthal velocities in the plane of the sky.

An interesting comparison object is the GG Tau circumbinary disk. A submillimetre continuum accretion stream[28] is seen to cross the gap that surrounds the central binary in GG Tau, with indications of shocked infrared molecular gas in the inner disk[29]. The angular radii of the rings in GG Tau and in HD 142527 are very similar, as are the morphologies of the GG Tau streamer and the eastern filament (although it is fainter relative to the outer disk in HD 142527). However, the GG Tau binary has a mass ratio of about 1, a separation of 0.25 arcsec and is aligned along the streamer, whereas in HD 142527 no stellar companion has been detected (see Supplementary Information for limits; mass ratios are >10 at 0.088 arcsec). The putative companions responsible for the streams in HD 142527 are of much lower mass than in GG Tau.

We performed high-contrast infrared imaging to attempt the detection of the possible accreting protoplanets that would be expected if the gap-crossing bridges observed in HCO^+ are indeed planet-induced gap-crossing accretion flows. Neglecting extinction, we could virtually rule out any companion more than about four times as massive as Jupiter, at a separation of 0.3–2.5 arcsec (Supplementary Information, section 4, and Supplementary Fig. 13, which also concerns the lack of close stellar companions). However, according to the hydrodynamical simulations the channelling protoplanets should be located inside the gap-crossing bridges. Our estimates for the hydrogen column number density, N_{H}, along the bridges correspond to a broad range of high visual ($\sim$5,500 Å) extinction values, 5 mag $\lesssim A_{\mathrm{V}} \lesssim$ 50 mag, for standard dust abundances. Any protoplanets embedded inside the bridges, and certainly those embedded in the dense horseshoe structure, will be obscured, and our mass limits will be correspondingly increased.

As well as the diffuse gas inside the gap and the gap-crossing filaments, a third feature of our observations is the horseshoe shape of the continuum, which was seen previously by the Submillimeter Array[8], Hawaii, but at coarser resolutions, and whose origin is still unclear. The millimetre continuum traces the total dust mass, so the north–south specific intensity ratio of 28 ± 0.5 reflects the underlying dust mass asymmetry. At its peak, the continuum may even be optically thick, as it coincides with a decrease in the HCO^+ emission. For a constant gas-to-dust mass ratio, such horseshoe-shaped mass asymmetries arise in models of planet-induced dynamical clearing. In general, these horseshoes can be produced by Rossby-wave instabilities, which are seen in high-resolution, three-dimensional simulations at the edge of sharp density gradients (P. Varnière, personal communication). However, horseshoes have also been modelled in the context of large-scale vortices induced by sharp viscosity gradients[30].

Another interpretation of the horseshoe continuum is that it results from a varying dust-to-gas ratio and azimuthal grain-size segregation. By contrast to the continuum, the outer disk is seen as a whole ring in HCO^+ (Supplementary Fig. 7), which is a tracer of dense gas, and is probably optically thick along the ring. A rarefaction of millimetre-sized dust grains to the south could perhaps explain the lack of a submillimetre continuum. Only small dust grains would be found in the south. These small grains efficiently scatter the near-infrared light seen in Fig. 1c, filling the opening of the horseshoe. However, azimuthal dust segregation has been predicted for centimetre-sized grains at co-rotation in two-fluid simulations of gap clearing by giant-planet formation[2], whereas millimetre-sized grains remain relatively unaffected. We explain in Supplementary Information why differential stellar heating cannot account for the observed north–south contrast.

The filamentary flows and the residual gas inside the gap are in qualitative agreement with planet-formation feedback on the parent disk, which feedback carves a gap in the dust distribution while still feeding stellar accretion through gap-crossing accretion streams. As detailed in Supplementary Information, the observed inflow velocity, together with the critical density of the molecular tracer and the section of the filaments, provide a lower bound to the mass inflow rate of $7\times10^{-9}M_{\odot}\ \mathrm{yr}^{-1}$. An upper bound of $2\times10^{-7}M_{\odot}\ \mathrm{yr}^{-1}$ can be estimated from the continuum mass in the filaments and their kinematic timescale. These estimates for the mass inflow rate are close to the observed stellar accretion rate[10], of $7\times10^{-8}M_{\odot}\ \mathrm{yr}^{-1}$, giving quantitative support to our suggestion that the HCO^+ filaments are inflows.

Received 12 August; accepted 31 October 2012.

Published online 2 January; corrected online 9 January 2013 (see full-text HTML version for details).

1. Lubow, S. H. & D'Angelo, G. Gas flow across gaps in protoplanetary disks. *Astrophys. J.* **641,** 526–533 (2006).
2. Fouchet, L., Gonzalez, J.-F. & Maddison, S. T. Planet gaps in the dust layer of 3D protoplanetary disks. I. Hydrodynamical simulations of T Tauri disks. *Astron. Astrophys.* **518,** A16 (2010).
3. Ayliffe, B. A., Laibe, G., Price, D. J. & Bate, M. R. On the accumulation of planetesimals near disc gaps created by protoplanets. *Mon. Not. R. Astron. Soc.* **423,** 1450–1462 (2012).
4. Zhu, Z., Nelson, R. P., Hartmann, L., Espaillat, C. & Calvet, N. Transitional and pretransitional disks: gap opening by multiple planets? *Astrophys. J.* **729,** 47–58 (2011).
5. van Boekel, R. *et al.* The building blocks of planets within the 'terrestrial' region of protoplanetary disks. *Nature* **432,** 479–482 (2004).
6. Fukagawa, M. *et al.* Near-infrared images of protoplanetary disk surrounding HD 142527. *Astrophys. J.* **636,** L153–L156 (2006).
7. Casassus, S. *et al.* The dynamically disrupted gap in HD 142527. *Astrophys. J.* **754,** L31–L35 (2012).
8. Ohashi, N. Observational signature of planet formation: the ALMA view. *Astrophys. Space Sci.* **313,** 101–107 (2008).
9. Öberg, K. I. *et al.* Disk imaging survey of chemistry with SMA. II. Southern sky protoplanetary disk data and full sample statistics. *Astrophys. J.* **734,** 98–109 (2011).
10. Garcia Lopez, R., Natta, A., Testi, L. & Habart, E. Accretion rates in Herbig Ae stars. *Astron. Astrophys.* **459,** 837–842 (2006).
11. Verhoeff, A. P. *et al.* The complex circumstellar environment of HD 142527. *Astron. Astrophys.* **528,** A91–A103 (2011).
12. Dodson-Robinson, S. E. & Salyk, C. Transitional disks as signposts of young, multiplanet systems. *Astrophys. J.* **738,** 131–145 (2011).
13. Fujiwara, H. *et al.* The asymmetric thermal emission of the protoplanetary disk surrounding HD 142527 seen by Subaru/COMICS. *Astrophys. J.* **644,** L133–L136 (2006).
14. Lyo, A.-R., Ohashi, N., Qi, C., Wilner, D. J. & Su, Y.-N. Millimeter observations of the transition disk around HD 135344B (SAO 206462). *Astron. J.* **142,** 151–160 (2011).
15. Mathews, G. S., Williams, J. P. & Ménard, F. 880 μm imaging of a transitional disk in Upper Scorpius: holdover from the era of giant planet formation? *Astrophys. J.* **753,** 59–70 (2012).
16. Tatulli, E. *et al.* Constraining the wind launching region in Herbig Ae stars: AMBER/VLTI spectroscopy of HD 104237. *Astron. Astrophys.* **464,** 55–58 (2007).
17. Kraus, S. *et al.* The origin of hydrogen line emission for five Herbig Ae/Be stars spatially resolved by VLTI/AMBER spectro-interferometry. *Astron. Astrophys.* **489,** 1157–1173 (2008).
18. Eisner, J. A. *et al.* Spatially and spectrally resolved hydrogen gas within 0.1 AU of T Tauri and Herbig Ae/Be Stars. *Astrophys. J.* **718,** 774–794 (2010).
19. Carr, J. S., Mathieu, R. D. & Najita, J. R. Evidence for residual material in accretion disk gaps: CO fundamental emission from the T Tauri spectroscopic binary DQ Tauri. *Astrophys. J.* **551,** 454–460 (2001).
20. Najita, J., Carr, J. S. & Mathieu, R. D. Gas in the terrestrial planet region of disks: CO fundamental emission from T Tauri Stars. *Astrophys. J.* **589,** 931–952 (2003).
21. Acke, B. & van den Ancker, M. E. Resolving the disk rotation of HD 97048 and HD 100546 in the [O I] 6300 Å line: evidence for a giant planet orbiting HD 100546. *Astron. Astrophys.* **449,** 267–279 (2006).
22. van der Plas, G. *et al.* The structure of protoplanetary disks surrounding three young intermediate mass stars. I. Resolving the disk rotation in the [OI] 6300 Å line. *Astron. Astrophys.* **485,** 487–495 (2008).
23. Salyk, C., Blake, G. A., Boogert, A. C. A. & Brown, J. M. High-resolution 5 μm spectroscopy of transitional disks. *Astrophys. J.* **699,** 330–347 (2009).
24. Pontoppidan, K. M. *et al.* Spectroastrometric imaging of molecular gas within protoplanetary disk gaps. *Astrophys. J.* **684,** 1323–1329 (2008).
25. van der Plas, G. *et al.* Evidence for CO depletion in the inner regions of gas-rich protoplanetary disks. *Astron. Astrophys.* **500,** 1137–1141 (2009).
26. Pontoppidan, K. M., Blake, G. A. & Smette, A. The structure and dynamics of molecular gas in planet-forming zones: a CRIRES spectro-astrometric survey. *Astrophys. J.* **733,** 84–100 (2011).
27. Sacco, G. G. *et al.* High-resolution Spectroscopy of Ne II emission from young stellar objects. *Astrophys. J.* **747,** 142 (2012).
28. Piétu, V., Gueth, F., Hily-Blant, P., Schuster, K.-F. & Pety, J. High resolution imaging of the GG Tauri system at 267 GHz. *Astron. Astrophys.* **528,** A81–A95 (2011).

29. Beck, T. L. *et al.* Circumbinary gas accretion onto a central binary: infrared molecular hydrogen emission from GG Tau A. *Astrophys. J.* **754,** 72–77 (2012).
30. Regály, Z., Juhász, A., Sándor, Z. & Dullemond, C. P. Possible planet-forming regions on submillimetre images. *Mon. Not. R. Astron. Soc.* **419,** 1701–1712 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This paper makes use of the following ALMA data: ADS/JAO.ALMA#2011.0.00465.S. ALMA is a partnership of the ESO, NSF, NINS, NRC, NSC and ASIAA. The Joint ALMA Observatory is operated by the ESO, AUI/NRAO and NAOJ. This work was also based on observations obtained at the Gemini Observatory. Financial support was provided by Millennium Nucleus P10-022-F (Chilean Ministry of Economy) and additionally by grant FONDECYT 1100221 and grant 284405 from the European Union FP7 programme.

Author Contributions General design of ALMA project, data analysis and write-up: S.C. Discussion of infrared observations of gas in cavities: G.v.d.P. Hydrodynamical modelling: S.P.M. ALMA data reduction: A.H. and E.F. Infrared-image processing: D.M., J.H. and J.H.G. Contributions to ALMA Cycle 0 proposal: A.J., F.M., D.W. and A.M.H. Design of ALMA observations: A.W., A.H. and S.C. Authors W.R.F.D. to A.W. contributed equally. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.C. (scasassus@u.uchile.cl).

LETTER

doi:10.1038/nature11781

How insulin engages its primary binding site on the insulin receptor

John G. Menting[1]*, Jonathan Whittaker[2]*, Mai B. Margetts[1], Linda J. Whittaker[2], Geoffrey K.-W. Kong[1], Brian J. Smith[1,3], Christopher J. Watson[4], Lenka Žáková[5], Emília Kletvíková[5], Jiří Jiráček[5], Shu Jin Chan[6], Donald F. Steiner[6], Guy G. Dodson[4]‡, Andrzej M. Brzozowski[4], Michael A. Weiss[2], Colin W. Ward[1] & Michael C. Lawrence[1,7]

Insulin receptor signalling has a central role in mammalian biology, regulating cellular metabolism, growth, division, differentiation and survival[1,2]. Insulin resistance contributes to the pathogenesis of type 2 diabetes mellitus and the onset of Alzheimer's disease[3]; aberrant signalling occurs in diverse cancers, exacerbated by cross-talk with the homologous type 1 insulin-like growth factor receptor (IGF1R)[4]. Despite more than three decades of investigation, the three-dimensional structure of the insulin–insulin receptor complex has proved elusive, confounded by the complexity of producing the receptor protein. Here we present the first view, to our knowledge, of the interaction of insulin with its primary binding site on the insulin receptor, on the basis of four crystal structures of insulin bound to truncated insulin receptor constructs. The direct interaction of insulin with the first leucine-rich-repeat domain (L1) of insulin receptor is seen to be sparse, the hormone instead engaging the insulin receptor carboxy-terminal α-chain (αCT) segment, which is itself remodelled on the face of L1 upon insulin binding. Contact between insulin and L1 is restricted to insulin B-chain residues. The αCT segment displaces the B-chain C-terminal β-strand away from the hormone core, revealing the mechanism of a long-proposed conformational switch in insulin upon receptor engagement. This mode of hormone–receptor recognition is novel within the broader family of receptor tyrosine kinases[5]. We support these findings by photo-crosslinking data that place the suggested interactions into the context of the holoreceptor and by isothermal titration calorimetry data that dissect the hormone–insulin receptor interface. Together, our findings provide an explanation for a wealth of biochemical data from the insulin receptor and IGF1R systems relevant to the design of therapeutic insulin analogues.

Insulin comprises two chains (A and B) containing three α-helices (residues A1–A8, A12–A18 and B9–B19) constrained by one intra- and two interchain disulphide bonds[6] (Fig. 1a). The insulin receptor is a disulphide-linked $(\alpha\beta)_2$ homodimer; the extracellular portion of each αβ protomer contains six domains (L1, CR, L2, FnIII-1, FnIII-2 and FnIII-3) and an insert domain (ID) within FnIII-2[7] (Fig. 1b). The α-chain component of the ID is terminated by a segment termed αCT, spanning residues 704–719 (using the numbering of the insulin receptor exon 11− isoform[8]). Two surfaces of insulin are understood to interact with the insulin receptor[9,10]. The first consists predominantly of hormone-dimerizing residues and contacts the primary binding site on the receptor (site 1; dissociation constant (K_d) ~6.4 nM) comprising the αCT segment from one insulin receptor α-chain and the central β-sheet of L1 (L1–β_2) of the other α-chain within the insulin receptor dimer[11–14] (Supplementary Fig. 1a). The second consists predominantly of hormone-hexamerizing residues and is proposed to interact with a secondary insulin receptor site (site 2; K_d ~400 nM) at the junction of FnIII-1 and FnIII-2 of the insulin receptor α-chain opposite to that contributing L1 to site 1 (Supplementary Fig. 1a)[7,9–11,13–15]. Structural study of the insulin receptor and IGF1R ectodomains is difficult; their large size, flexible multidomain morphology and extensive glycosylation and disulphide bonding posing challenges in production and crystallization. Only three ectodomain structures exist: those of the *apo*-insulin receptor[7,12] and the L1–CR–L2 fragments of insulin receptor[16] and IGF1R[17] (3.7 Å, 2.3 Å and 2.6 Å resolution, respectively); none is in complex with ligand.

We present here the first view of the insulin–site 1 interaction on the basis of four crystal structures (complexes A, B, C and D) varyingly using the L1–CR construct IR310.T (Fig. 1c) and the L1–CR–L2–(FnIII-1)–αCT(704–719) homodimeric construct IR593.αCT (Fig. 1d), which has αCT(704–719) linked to FnIII-1. Complex A (3.9 Å resolution) contains human-insulin-bound IR310.T in complex with a Fab fragment from monoclonal antibody 83-7 (ref. 7) and exogenous αCT(704–719) peptide. Complex B (4.3 Å resolution) is the same as complex A but uses the high-affinity, truncated insulin analogue [D-Pro^{B26}]-DTI-NH_2[18]. Complex C (4.4 Å resolution) is the same as complex B but uses the longer peptide αCT(697–719). Complex D (4.4 Å resolution) contains bovine insulin bound to IR593.αCT in complex with a Fab fragment from monoclonal antibody 83-14 (ref. 7).

The limited diffraction of these crystals necessitated the use of the following: (1) weak data at the diffraction limit (a process validated by the half-data set Pearson correlation coefficient $CC_{1/2}$ criterion[19,20] and ultimately by the evident quality (see below) of the resultant electron-density difference maps (EDDMs)); (2) B-factor-sharpened EDDMs in model building[21]; and (3) dynamic 'reference-structure' and 'jelly-body' restraints during individual atomic-coordinate refinement whereby we exploited the known higher-resolution structures of the underlying fragments[22,23]. Molecular replacement located the insulin receptor and Fab domains and generated a clear solution for the insulin three-helix 'core' structure within each complex, positioning the core identically on the respective L1–β_2 surfaces. The strongest feature in the resultant EDDMs was a common, helical-like structure lying adjacent to the insulin B-helix on the L1–β_2 surface that, upon B-factor sharpening, exhibited side-chain density of distinctive sizes and at spacing consistent with an α-helix. In complex D, this feature was connected to the C-terminal strand of FnIII-1, allowing its unambiguous interpretation as αCT and restricting its prima facie sequence register to an ambiguity of at most one residue unit, with one alternative strongly favoured. We confirmed our register assignment objectively by devising a broad-window analysis of sequence fit both to the higher-resolution EDDM of complex A and to the profiled residue environments within that complex. Crystallographic processing and refinement statistics are provided in Table 1 and Supplementary Table 1, with further detail in Methods.

[1]Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia. [2]Department of Biochemistry, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106, USA. [3]Department of Chemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia. [4]York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK. [5]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., 16610 Prague, Czech Republic. [6]Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA. [7]Department of Medical Biology, University of Melbourne, Royal Parade, Parkville, Victoria 3010, Australia.
*These authors contributed equally to this work.
‡Deceased.

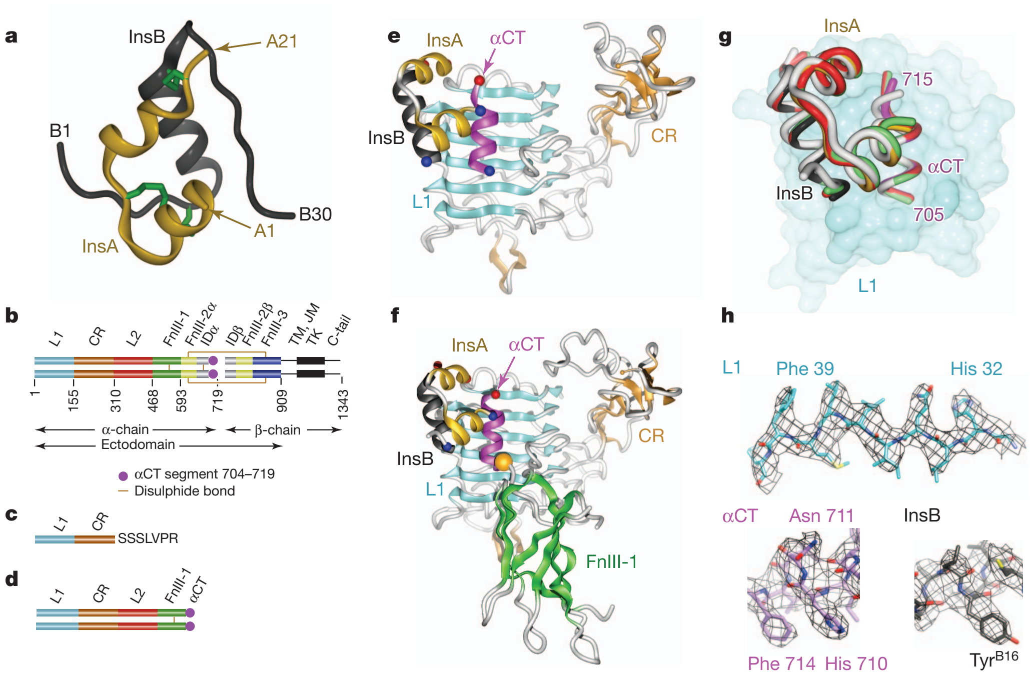


Figure 1 | Structure of insulin, insulin receptor and the site 1 complexes. **a**, Insulin. InsA, A chain; InsB, B chain. **b**, Insulin receptor. CR, Cys-rich domain; FnIII-1, FnIII-2, FnIII-3, first, second and third fibronectin type III domains; ID, insert domain; L1, L2, first and second leucine-rich-repeat domains; TK, tyrosine kinase; TM, JM, transmembrane and juxtamembrane segments. **c**, **d**, IR310.T and IR593.αCT domain structure, respectively. **e**, **f**, Insulin-bound site 1 in complexes A and D, respectively. Blue, red spheres, observed chain termini; orange sphere, FnIII-1–αCT junction. **g**, Overlay of insulin-bound site 1 in complexes A (coloured as in **e**), B (red), C (green) and D (white). **h**, Sample $2F_{obs} - F_{calc}$ map volumes ($B_{sharp} = -160$ Å^2; contours = 1.1–1.5σ) for complex A.

The structures of the insulin-bound site 1 assembly within complexes A and D are shown in Fig. 1e, f, respectively; those within complexes B and C are effectively isomorphous with complex A (Fig. 1g). The Fab-complexed forms of the IR310.T and IR593.αCT monomers are shown in Supplementary Fig. 2a, b, respectively. As prepared and crystallized, complex D contains a dimer of IR593.αCT homodimers crosslinked by four insulins; further description of this assembly is provided in Supplementary Fig. 2c–e. Despite the non-native connection of αCT to FnIII-1 in IR593.αCT (Fig. 1d), the insulin-bound site 1 assembly within complex D is also superimposable on that of complex A (Fig. 1g). Refinement of these complexes provided EDDMs that revealed side-chain bulk for all components within the observed site 1 interface (Fig. 1h). The mode of insulin–site 1 interaction emerging from these structures is as follows. The insulin B-helix (B7–B21) engages the C-terminal end of the L1–β_2 strands; the insulin A-chain has no interaction with L1 (Fig. 1e, f). Both chains interact extensively with αCT. Notably, the αCT helix is, with respect to its *apo*-insulin receptor counterpart, both displaced on the L1–β_2 surface and C-terminally extended to include residues 711–715 (see Fig. 2a and further discussion below). Residues B22–B30 are not resolved in these structures, but, critically, the αCT helix occupies volume that would otherwise contain insulin residues B26–B30 if the latter retained their receptor-free conformation (see Fig. 2b and Supplementary Fig. 3a). Displacement of residues B26–B30 from the insulin core confirms a long-speculated[24] induced fit upon insulin receptor binding (Supplementary Fig. 1b). In contrast, the helical core of insulin seems, as far as can be discerned at the described resolution, closely similar to that of uncomplexed hormone (Supplementary Fig. 3b). The models suggest that the two most critical hormone-engaging residues in αCT are (1) His 710, which inserts into a pocket formed by invariant[10] insulin residues Val^{A3}, Gly^{B8}, Ser^{B9} and Val^{B12}; and (2) Phe 714, which occupies a hydrophobic crevice formed by invariant insulin residues

Table 1 | X-ray diffraction data collection statistics

	Complex A, data set 1	Complex A, data set 2	Complex B	Complex C	Complex D
	Human insulin + IR310.T–83-7 + αCT(704–719)	Human insulin + IR310.T–83-7 + αCT(704–719)	[D-Pro^{B26}]-DTI-NH_2 + IR310.T–83-7 + αCT(704–719)	[D-Pro^{B26}]-DTI-NH_2 + IR310.T–83-7 + αCT(697–719)	Bovine insulin + IR593.αCT/83-14
Space group	*P*23	*P*23	*P*23	*P*23	*I*2
Cell dimensions					
a, *b*, *c* (Å)	168.15, 168.15, 168.15	168.91, 168.91, 168.91	169.49, 169.49, 169.49	169.23, 169.23, 169.23	118.15, 140.10, 190.02
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 95.04, 90
Completeness (%)	99.8 (99.8)*	98.3 (99.5)	99.7 (100.0)	98.5 (98.1)	87.7 (84.9)†
Resolution (Å)	59.5–4.0 (4.1–4.0)	46.8–3.9 (4.0–3.9)	29.6–4.3 (4.53–4.3)	29.5–4.30 (4.54–4.3)	56.3–4.4 (4.5–4.4)
R_{merge}	0.156 (2.36)	0.089 (1.57)	0.188 (2.69)	0.101 (1.09)	0.206 (1.31)
‡R_{pim}	0.039 (0.57)	0.048 (0.83)	0.067 (0.97)	0.053 (0.59)	0.143 (0.89)
$I/\sigma I$	11.5 (1.6)	9.0 (1.0)	7.8 (1.0)	8.2 (1.4)	2.8 (0.8)
Redundancy	17.3 (17.4)	4.1 (4.2)	8.8 (8.8)	4.4 (4.3)	2.9 (2.8)
§$CC_{1/2}$	0.998 (0.580)	0.998 (0.342)	0.996 (0.330)	0.997 (0.438)	0.980 (0.232)

Each data set was collected from a single crystal.
* Highest-resolution shell statistics are shown in parentheses.
† Data completeness was compromised by radiation damage.
‡ R_{pim}, precision indicating merging *R*-factor[30].
§ $CC_{1/2}$, Pearson correlation coefficient between independently merged halves of the data set[19]. Highest-resolution shell $CC_{1/2}$ values are significant at at least $P = 0.001$.

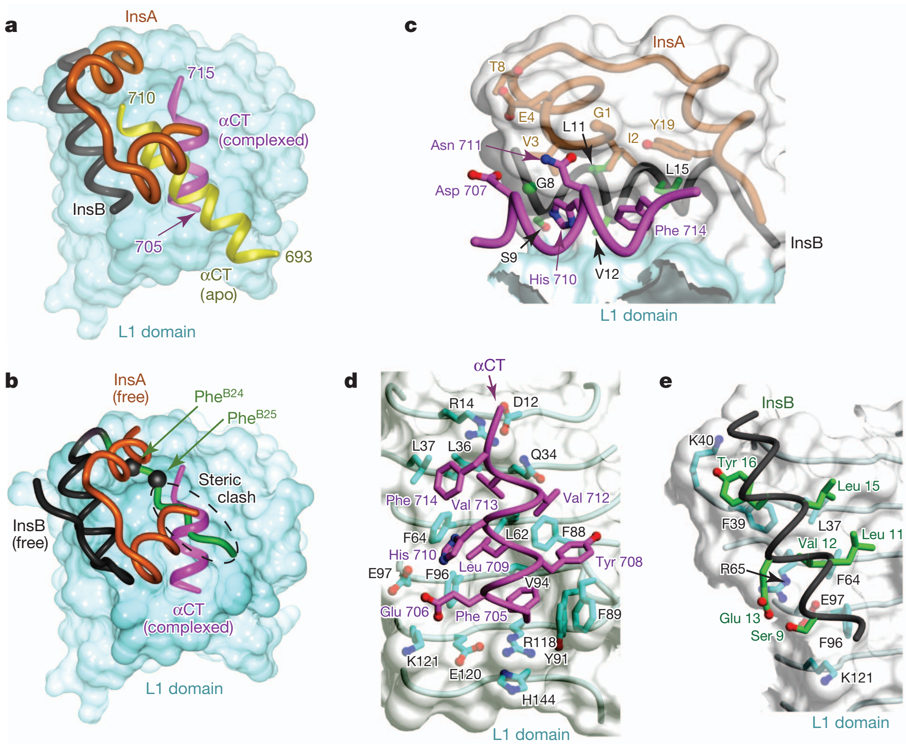


Figure 2 | The insulin–site 1 interaction. **a**, Altered disposition of αCT with respect to that in *apo*-insulin receptor[12]. **b**, Superposition (via the A- and B-chain helices) of receptor-free insulin onto the insulin–site 1 complex, indicating steric clash of the B-chain C-terminal segment (green) with αCT. See also Supplementary Fig. 3a. **c**, Interaction (at the bulk side-chain level of detail) between αCT(704–719) and insulin. View direction is parallel to L1–β2 surface (cyan); white shading shows insulin surface. **d**, **e**, Interaction (at the bulk side-chain level of detail) between L1 and αCT and between L1 and insulin B-chain helix, respectively. Panels based on complex A.

Gly^{A1}, Ile^{A2}, Tyr^{A19}, Leu^{B11}, Val^{B12} and Leu^{B15} (Fig. 2c). Asn 711 is directed towards Gly^{A1}, Val^{A3} and Glu^{A4}. The hydrophobic face of αCT comprising Phe 705, Tyr 708, Leu 709, Val 712 and Val 713 engages a non-polar groove on L1-β_2 formed by Leu 36, Leu 37, Leu 62, Phe 64, Phe 88, Phe 89, Val 94 and Phe 96 (Fig. 2d). Direct interaction between the insulin B-chain B9–B19 helix and L1 seems to be limited: Val^{B12} is positioned between Phe 39, Phe 64 and Arg 65, whereas Tyr^{B16} adjoins Phe 39 (Fig. 2e). Detailed rotameric and atomic level detail of the above interactions is precluded by the limited diffraction.

Relocation of αCT upon hormone binding is supported by recent studies indicating that residues buried beneath αCT in the *apo*-ectodomain may directly contact insulin[25], and that peptide substitutions in the αCT segment upstream of Phe 705 that increase the affinity of the segment for L1 concomitantly impair insulin binding[12]. To investigate further, we used complex C to test whether the relocation observed in complexes A, B and D is a consequence of αCT truncation, as in the *apo*-insulin receptor the visible interaction of αCT with L1–β_2 begins at Glu 697. It is not—the relocation persists in complex C (Fig. 1g) with no detectable ordering of residues 697–704.

We validated our interpretation of the αCT engagement with L1–β_2 and insulin further by isothermal titration calorimetry (ITC). First, Ala-substituted αCT(704–719) peptides were titrated against the L1–CR–L2 construct IR485 (ref. 16). Second, insulin was titrated against IR485 pre-saturated with Ala-substituted αCT(704–719) (Supplementary Tables 2 and 3 and Supplementary Fig. 4). Individual F705A and L709A substitutions significantly impaired αCT binding to IR485, supporting their modelled engagement with L1–β_2. In contrast, individual H710A, N711A and F714A substitutions significantly impaired insulin binding, supporting their modelled engagement with insulin. These data are summarized schematically in Fig. 3a. An apparent exception is V713A, which affects insulin but not αCT affinity, despite Val 713 being modelled as engaged with L1–β_2. However, this

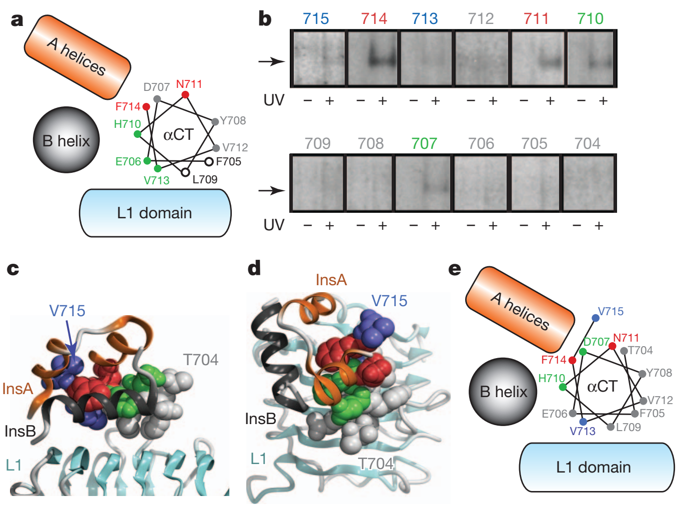


Figure 3 | Insulin interactions in L1–CR–L2 mini-insulin receptor and *holo*-insulin receptor. **a**, Helical wheel representation of ITC-derived insulin affinities for IR485 in the presence of Ala-substituted αCT(704–719) (red, >100× reduction upon Ala substitution; green, >10× reduction; grey, <10× reduction; open circle, not determined). **b**, Reducing gel autoradiograms obtained from *holo*-insulin receptor after photo-crosslinking of αCT helix to bound ^{125}I-[Tyr^{A14}]-insulin. Arrowed band indicates crosslinked insulin receptor α-chain–insulin A chain. Colours indicate crosslinking efficiency (red, strong; green, medium; blue, weak; grey, none). **c**, **d**, Qualitative crosslinking efficiency from **b** mapped onto αCT segment within the site 1 complexes. **e**, Helical wheel representation of crosslinking data presented in **b**–**d**.

observation aligns with the disorder of Val 713 in the *apo*-insulin receptor ectodomain structure[12]; the interaction of Val 713 with L1–β_2 may thus instead be required to stabilize the critical[11] insulin-engaging residue Phe 714 at the terminus of the αCT helix.

The insulin–site 1 interactions described here are further supported by a 'photo-scan' of the *holo*-insulin receptor αCT segment, wherein individual αCT residues were successively substituted by the photo-activatable residue *para*-azido-Phe (Pap) and then exposed to ultraviolet light to probe covalent interaction with [^{125}I]-Tyr^{A14} insulin[25]. Strong crosslinking was observed for photo-probes at sites 711 and 714, medium at sites 707 and 710, and weak at sites 713 and 715; no crosslinking was observed at sites 705, 706, 708, 709 or 712 (Fig. 3b). This photo-crosslinking efficiency pattern in *holo*-insulin receptor exhibits a marked correlation with and supports the insulin–site 1 juxtaposition described above (Fig. 3c–e).

Complexes B and C probed whether the stable β-turn at B24–B26 in receptor-free [D-Pro^{B26}]-DTI-NH_2[18] might result in ordering of residues B22–B26 upon binding site 1—it did not, as these residues remained unresolved. We investigated the insulin B-chain C-terminal strand further by seeking to photo-crosslink individual Pap probes at B24 and B25 to the truncated L1–CR–L2 construct IR468[26] in the presence of an αCT peptide. In both *holo*- and ectodomain insulin receptor complexes, these probes respectively crosslink to L1 and αCT[27,28]. We observed photo-crosslinking for Pap^{A3}, Pap^{A8} and Pap^{B16} (as in *holo*-insulin receptor[27]) but not for Pap^{B24} or Pap^{B25} (Supplementary Fig. 5). We thus propose that full stabilization of B22–B30 upon detachment from the hormone core requires the presence of the FnIII domains.

Superposition of the site 1 complexed assembly onto the L1 domain in the *apo*-insulin receptor ectodomain structure (Supplementary Fig. 6) shows that the insulin surface proposed[10] to bind site 2 overlays the FnIII-1–FnIII-2 junction of the alternative αβ protomer, that is, the proposed site 2 itself. However, significant steric clash is present, indicating that insulin bridging of sites 1 and 2 requires relative displacement of the FnIII-1 and FnIII-2 domains away from the L1–CR module. Such change would alter the relative location of the membrane-proximal FnIII-3 domains, potentially initiating signal transduction[13].

Concordance of our structures with the extensive biochemical literature of this system is presented in Supplementary Discussion. Briefly, the structures are consistent with alanine scanning mutagenesis of the insulin receptor L1 domain, revealing now that of the 'hotspot' residues, the set Gln 34, Leu 36, Leu 37, Phe 64, Phe 88, Phe 89 and Glu 120 contact the αCT segment whereas only Phe 39 contacts insulin. The structural engagement of insulin receptor residues His 710 and Phe 714 with insulin concords with (1) alanine scanning mutagenesis of the insulin receptor ectodomain; (2) hydrogen/deuterium exchange studies of the closely related type 1 insulin-like growth factor receptor (IGF1R); and (3) the lack of hormone binding by insulin receptor chimaeras and IGF1R chimaeras containing the αCT segment from the insulin-receptor-related receptor, which has threonine at the residue equivalent to insulin receptor Phe 714 (IGF1R Phe 701). Our structures are also consistent with mutational analysis of insulin, showing now that the hotspot residues Gly^{A1}, Ile^{A2}, Val^{A3}, Tyr^{A19}, Gly^{B8} and Leu^{B11} make contact with the key αCT residues His 710 and Phe 714 (Fig. 2c). Although residues Gly^{B23}, Phe^{B24} and Phe^{B25} are unresolved here, evidence is presented for the likely repositioning of the B-chain C terminus into the volume between the αCT segment, the L1–β_2 surface and the adjacent CR domain; such repositioning rationalizes the enhanced affinity of analogues containing D-amino-acid substitutions at B24.

Our structural and biochemical findings provide groundbreaking insight into the primary insulin–insulin receptor interaction. Particularly important in binding is the αCT segment, which is without known counterpart in other mammalian receptor tyrosine kinases[5], although it is conserved among metazoan insulin receptor and IGF1R. The concomitant structural re-arrangements of hormone and receptor in our model rationalize long-standing anomalies in clinical insulin variants associated with diabetes[29] and define new structural targets for novel therapeutics targeting these receptors.

METHODS SUMMARY

IR310.T was produced by proteolysis of an engineered mini-receptor precursor cIR485 expressed from a stable Chinese hamster ovary Lec8 cell line and purified by ion-exchange and size-exclusion chromatography (SEC), followed by combination with 83-7, re-purification by SEC and finally combination with excess αCT peptide and insulin (or analogue). Crystallization conditions for IR310.T-derived complexes A, B and C were 0.9–1.1 M tri-sodium citrate, 0.1 M imidazole-HCl (pH 8.0) + 0.02% sodium azide (NaN_3). IR593.αCT was expressed from a stable Chinese hamster ovary Lec8 cell line followed by insulin affinity chromatography and then SEC of the resultant bovine insulin complex. Samples were combined with 83-14 and re-purified by SEC. The crystallization condition for the IR593.αCT-derived complex D was 9% polyethylene glycol 3350, 200 mM proline, 100 mM HEPES-NaOH (pH 7.5). Diffraction data were collected at the Australian Synchrotron (beamline MX2) and Diamond Light Source (DLS; beamline I24). Structures were solved by molecular replacement. Introduction of Pap residues into the αCT segment used orthogonal tRNA/amber suppression technology. Mutant holoreceptors were expressed transiently using co-transfected 293PEAK rapid cells in Pap-containing medium. Cell detergent lysates were then subjected to wheat-germ agglutinin chromatography; eluates containing the receptor were concentrated before photo-crosslinking to ^{125}I-[Tyr^{A14}] insulin. Photo-crosslinked products were resolved by gel electrophoresis and auto-radiographed. IR468 was produced by transient expression in HEK293 cells. ^{125}I-photo-activatable insulin analogues were incubated with IR468 in the presence of 10^{-5} M αCT(703–719)–Myc and then photo-crosslinked. Samples were resolved by gel electrophoresis and auto-radiographed. For ITC, Ala-substituted αCT peptides (80 μM concentration) were titrated into a sample cell containing IR485 (10 μM concentration); and zinc-free porcine insulin (32–48 μM concentration) was titrated into a sample cell containing IR485 (4–6 μM concentration) pre-incubated with Ala-substituted αCT peptide (10× molar concentration). All samples were in Tris-buffered saline (pH 8.0) plus azide (TBSA).

Full Methods and any associated references are available in the online version of the paper.

Received 25 July; accepted 12 November 2012.

1. Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into insulin action. *Nature Rev. Mol. Cell Biol.* **7,** 85–96 (2006).
2. Cohen, P. The twentieth century struggle to decipher insulin signalling. *Nature Rev. Mol. Cell Biol.* **7,** 867–873 (2006).
3. Talbot, K. *et al.* Demonstrated brain insulin resistance in Alzheimer's disease patients is associated with IGF-1 resistance, IRS-1 dysregulation, and cognitive decline. *J. Clin. Invest.* **122,** 1316–1338 (2012).
4. Pollak, M. The insulin and insulin-like growth factor receptor family in neoplasia: an update. *Nature Rev. Cancer* **12,** 159–169 (2012).
5. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117–1134 (2010).
6. Adams, M. J. *et al.* Structure of rhombohedral 2 zinc insulin crystals. *Nature* **224,** 491–495 (1969).
7. McKern, N. M. *et al.* Structure of the insulin receptor ectodomain reveals a folded-over conformation. *Nature* **443,** 218–221 (2006).
8. Seino, S. & Bell, G. I. Alternative splicing of human insulin receptor messenger RNA. *Biochem. Biophys. Res. Commun.* **159,** 312–316 (1989).
9. Schäffer, L. A model for insulin binding to the insulin receptor. *Eur. J. Biochem.* **221,** 1127–1132 (1994).
10. De Meyts, P. Insulin and its receptor: structure, function and evolution. *Bioessays* **26,** 1351–1362 (2004).
11. Ward, C. W. & Lawrence, M. C. Ligand-induced activation of the insulin receptor: a multi-step process involving structural changes in both the ligand and the receptor. *Bioessays* **31,** 422–434 (2009).
12. Smith, B. J. *et al.* Structural resolution of a tandem hormone-binding element in the insulin receptor and its implications for design of peptide agonists. *Proc. Natl Acad. Sci. USA* **107,** 6771–6776 (2010).
13. Ward, C. W. & Lawrence, M. C. Similar but different: ligand-induced activation of the insulin and epidermal growth factor receptor families. *Curr. Opin. Struct. Biol.* **22,** 360–366 (2012).
14. Kiselyov, V. V., Versteyhe, S., Gauguin, L. & De Meyts, P. Harmonic oscillator model of the insulin and IGF1 receptors' allosteric binding and activation. *Mol. Syst. Biol.* **5,** 243 (2009).
15. Whittaker, L., Hao, C., Fu, W. & Whittaker, J. High-affinity insulin binding: insulin interacts with two receptor ligand binding sites. *Biochemistry* **47,** 12900–12909 (2008).

16. Lou, M. *et al.* The first three domains of the insulin receptor differ structurally from the insulin-like growth factor 1 receptor in the regions governing ligand specificity. *Proc. Natl Acad. Sci. USA* **103,** 12429–12434 (2006).
17. Garrett, T. P. *et al.* Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. *Nature* **394,** 395–399 (1998).
18. Jiráček, J. *et al.* Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues. *Proc. Natl Acad. Sci. USA* **107,** 1966–1970 (2010).
19. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).
20. Evans, P. Resolving some old problems in protein crystallography. *Science* **336,** 986–987 (2012).
21. Brünger, A. T., DeLaBarre, B., Davies, J. M. & Weis, W. I. X-ray structure determination at low resolution. *Acta Crystallogr. D* **65,** 128–133 (2009).
22. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* **68,** 368–380 (2012).
23. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).
24. Hua, Q. X., Shoelson, S. E., Kochoyan, M. & Weiss, M. A. Receptor binding redefined by a structural switch in a mutant human insulin. *Nature* **354,** 238–241 (1991).
25. Whittaker, J. *et al.* α-Helical element at the hormone-binding surface of the insulin receptor functions as a signaling element to activate its tyrosine kinase. *Proc. Natl Acad. Sci. USA* **109,** 11166–11171 (2012).
26. Kristensen, C., Wiberg, F. C. & Andersen, A. S. Specificity of insulin and insulin-like growth factor I receptors investigated using chimeric mini-receptors. Role of C-terminal of receptor α subunit. *J. Biol. Chem.* **274,** 37351–37356 (1999).
27. Xu, B. *et al.* Diabetes-associated mutations in insulin: consecutive residues in the B chain contact distinct domains of the insulin receptor. *Biochemistry* **43,** 8356–8372 (2004).
28. Xu, B. *et al.* Decoding the cryptic active conformation of a protein by synthetic photo-scanning. Insulin inserts a detachable arm between receptor domains. *J. Biol. Chem.* **284,** 14597–14608 (2009).
29. Xu, B. *et al.* Diabetes-associated mutations in insulin identify invariant receptor contacts. *Diabetes* **53,** 1599–1602 (2004).
30. Weiss, M. S. Global indicators of X-ray data quality. *J. Appl. Cryst.* **34,** 130–135 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements This Letter is dedicated to our co-author, the late Guy Dodson, in recognition of his lifetime contribution to the study of the structure of insulin. This work was supported by Australian National Health and Medical Research Council (NHMRC) Project grants 516729, 575539 and 1005896 and the Hazel and Pip Appel Fund (to M.C.L.), NHMRC Independent Research Institutes Infrastructure Support Scheme Grant 361646 and Victorian State Government Operational Infrastructure Support Grant (to the Walter and Eliza Hall Institute of Medical Research), NIH grant no. DK40949 (to M.A.W. and J.W.) and American Diabetes Association grant no. 1-11INI-31 (to J.W.), Grant Agency of the Czech Republic grant P207/11/P430 (to L.Z.), Research Project of the Academy of Sciences of the Czech Republic RVO:61388963 (to the Institute of Organic Chemistry and Biochemistry), NIH grants DK13914 and DK20595 (to D.F.S.), a BBSRC PhD studentship (to C.J.W.) and the UoY Research Priming Fund (to the York Structural Biology Laboratory). Part of this research was undertaken on the MX2 beamline at the Australian Synchrotron (AS), Victoria, Australia. We thank the DLS for access to beamline I24 and the Australian International Synchrotron Access Program for travel funds. We thank P. Colman and J. Gulbis, our colleagues at CSIRO and the AS beam line staff for their support; J. Turkenburg for assistance in collecting data at DLS; K. Huang for assistance with midi-receptor photo-crosslinking; Q.-X. Hua and Y. Yang for discussion of NMR studies of insulin; S.-Q. Hu, S. H. Nakagawa, N. F. Phillips and S. Wang for assistance with insulin analogue synthesis; P. G. Katsoyannis for advice about the synthesis of photo-reactive insulin analogues and for providing an initial set of Pap analogues; K. Siddle for supplying the 83-7 and 83-14 hybridomas; L. Lu and the fermentation group CSIRO Materials Science and Engineering for large-scale cell culture.

Author Contributions J.G.M. and J.W. contributed equally to the paper. J.G.M. purified and crystallized samples, collected data and performed the ITC study; J.W. and L.J.W. performed receptor photo-crosslinking experiments; M.B.M. performed molecular biology, cell culture and crystallization experiments; S.J.C. performed insulin photo-crosslinking experiments; G.K.-W.K. and C.J.W. performed crystallography experiments; B.J.S. performed calculations; E.K., L.Z. and J.J. prepared insulin analogues; C.W.W., M.A.W., J.W., D.F.S., S.J.C., J.G.M. and M.C.L. designed the experiments and analysed data. C.W.W., M.A.W., A.M.B., G.G.D. and M.C.L. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates and structure factors for complexes A, B, C and D have been deposited with the Protein Data Bank under accession codes 3W11, 3W12, 3W13 and 3W14, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C.L. (lawrence@wehi.edu.au) or M.A.W. (michael.weiss@case.edu).

METHODS

Cloning, production and purification of IR310.T. High yield and correct folding of IR310.T was achieved by design of a thrombin-cleavable version of IR485[16] (termed cIR485) for use as precursor, followed by production and purification protocols for cIR485 similar to those described for IR485[16]. A 1,741-nucleotide base fragment of insulin receptor was synthesized (DNA2.0) and inserted into the HindIII–XbaI sites of the vector pEE14 (Lonza). The fragment encoded a protein cIR485 ('cleavable' IR485) which consists, in order, of (1) the 27-residue insulin receptor native signal sequence; (2) residues 1–310 of the insulin receptor α-chain (the L1–CR module); (3) a thrombin cleavage site SSSLVPRGSSS; (4) residues 311–485 of the insulin receptor α-chain (the L2 domain); (5) an enterokinase cleavage site DDDDK; and (6) a c-Myc purification tag EQKLISEEDLN[31] (see Supplementary Fig. 7). A 106-nucleotide-base non-coding fragment 5′-GTCG ACGGTACCCCGGGGAATTAATTCCGGGGGCCGCCTCGGAGCATGACC CCCGCGGGCCAGCGCCGCGCGCTCTGATCCGAGGAGACCCCGCGCTC CCGCAGCC-3′ was included between the HindIII site and the region expressing the signal peptide; bases 29–106 of this fragment correspond to those immediately upstream of the human insulin receptor coding region. CHO Lec8 cells (CRL-1737; ATCC)[32] were transfected with the cIR485-containing expression vector using Lipofectamine 2000 (Life Technologies). Selection was undertaken in the presence of methionine sulphoximine (MSX), with cIR485 production monitored via western blot with monoclonal antibodies 83-7[33] and 9E10[34]. Production scale-up, performed under contract by CSIRO, was in spinner flasks (New Brunswick Scientific) containing DMEM/F12 medium (Life Technologies) supplemented with 10% fetal bovine serum (FBS; Life Technologies) plus 25 μM MSX. The medium was supplemented with 0.8 mM butyric acid (Sigma-Aldrich) during the last week of the production phase. A total of 50 l of culture supernatant was collected; harvests were filtered (0.2 μm cutoff) and then concentrated 10-fold (30 kDa cutoff) after addition of 0.1 mM phenylmethanesulphonylfluoride (PMSF; Sigma-Aldrich) and 0.02% sodium azide (NaN_3). Accumulated supernatant harvests were adjusted to pH 8 by addition of 3 M Tris-HCl (pH 8.5). Initial purification of cIR485 was by 9E10 antibody affinity chromatography and size-exclusion chromatography (SEC) using procedures effectively identical to those described for purification of IR485[16]. Purified cIR485 in Tris-buffered saline (24.8 mM Tris-HCl pH 8.0, 137 mM NaCl, 2.7 mM KCl) plus 0.02% NaN_3 (TBSA) was then incubated overnight at 37 °C with 0.25 units human thrombin (Roche) per mg of cIR485 in the presence of 10 mM $CaCl_2$. Completion of proteolysis was assessed by SDS–PAGE, which revealed bands corresponding to the estimated molecular mass of the IR310.T fragment and the insulin receptor L2 domain (51 kDa and 30 kDa, respectively; Supplementary Fig. 8a). Western blotting with monoclonal antibody 83-7 confirmed that the CR domain was contained in the upper band alone. The sample was diluted 8-fold in buffer A (10 mM ethanolamine-HCl, pH 9.6 + 0.02% NaN_3), centrifuged for 5 min at 17,000*g* to remove particulates and then loaded onto a Mono Q 5/50 GL column (GE Healthcare). The sample was eluted with a 60 column-volume gradient of buffer A to buffer B (10 mM ethanolamine-HCl, 400 mM NaCl, pH 9.6 + 0.02% NaN_3; Supplementary Fig. 8b) and the fractions assessed by SDS–PAGE. Fractions containing IR310.T were pooled, concentrated and re-run in TBSA on a Superdex 200 10/300 column (GE Healthcare). The chromatogram exhibited three overlapping peaks, probably arising from multimerization (Supplementary Fig. 8c). SDS–PAGE of fractions revealed the presence of three closely spaced bands, which we attributed to varying glycosylation (Supplementary Fig. 8d). The final yield of IR301.T was ~0.5 mg ml^{-1} of cell culture.

Large-scale production and purification of Fab 83-7. A hybridomal cell line expressing monoclonal antibody 83-7[33] was a gift from K. Siddle. Large-scale production and purification of monoclonal antibody 83-7 and preparation of Fab 83-7 from monoclonal antibody 83-7 were based on protocols described previously[7]. In particular, monoclonal antibody 83-7 was proteolysed with dithiothreitol-activated papain (Sigma-Aldrich) at 37 °C. Proteolysis was stopped by adding iodoacetamide (IAA; Sigma-Aldrich) and the reaction mixture then passed down a Superdex 200 26/60 column (GE Healthcare). Fractions containing 83-7 $F(ab')_2$ were isolated and reduced with mercaptoethylamine (Sigma-Aldrich) and then alkylated with IAA, followed by further SEC and cation-exchange chromatography on Mono S (GE Healthcare).

Production of Fab 83-7-complexed IR310.T. IR310.T was mixed with a slight molar excess of Fab 83-7 and the complex purified by SEC using a Superdex S200 10/300 column (GE Healthcare) in TBSA buffer. Fractions containing the complex of IR310.T/Fab 83-7 were then pooled, concentrated and exchanged into 10 mM HEPES-NaOH (pH 7.5) + 0.02% NaN_3 for crystallization screens.

Cloning, production and purification of IR593.αCT. CHO Lec8 cells stably expressing IR593.αCT were obtained from CSIRO. These cells were derived by co-transfection with both the linearized expression plasmid pEE14 (Lonza) and a pEF puromycin plasmid (Springer Labs). The pEE14 expression plasmid contained, inserted between its HindIII–XbaI sites, a fragment encoding (in order) the 27-amino-acid residue insulin receptor signal sequence, insulin receptor residues 1–593 and insulin receptor residues 704–719 (see Supplementary Fig. 9). The 106-nucleotide-base non-coding fragment described above was also included here between the HindIII site and the start codon. Selection was undertaken in the presence of puromycin, with IR593.αCT production monitored via europium-labelled insulin binding assays[35]. Production scale-up, performed under contract by CSIRO, was in spinner flasks (New Brunswick Scientific) containing DMEM/F12 medium (Life Technologies) supplemented with 10% FBS plus 10 μg ml^{-1} puromycin. Accumulated supernatant batches were adjusted to pH 8 by addition of 3 M Tris-HCl (pH 8.5). PMSF was added to 0.1 mM concentration and NaN_3 to 0.02%. The culture supernatant was filtered (0.2 μm cutoff), concentrated 10-fold (10 kDa cutoff) and pumped through a Sephacryl S100 column (GE Healthcare) onto an in-house manufactured *bis*-Boc-insulin MiniLeak medium resin column (Kem-En-Tec)[36]. The latter was eluted with bovine insulin in TBSA. PMSF was again added to 0.1 mM concentration and the eluates concentrated for Superdex S200 26/60 or S200 10/300 (GE Healthcare) SEC. The SEC profiles showed a single major peak corresponding to the insulin-complexed IR593.αCT homodimer and a further peak corresponding to excess bovine insulin (Supplementary Fig. 8g). At higher concentrations, an additional peak appeared corresponding to a dimer of IR593.αCT homodimers (Supplementary Fig. 8h). Contents of fractions were analysed by SDS–PAGE (data not shown). The final yield of IR593.αCT was ~0.4 mg l^{-1} of cell culture supernatant.

Large-scale production and purification of Fab 83-14. A hybridomal cell line expressing Mab 83-14[33] was a gift from K. Siddle. Cells were grown in 950 cm^2 roller bottles at 37 °C in 5% CO_2 (balanced with air) at 10 r.p.m., using approximately 300 ml H-SFM medium supplemented with 5% fetal calf serum (FCS). Typically, 2–3 l of cell culture at a viable cell density of $2–3 \times 10^6$ cells ml^{-1} was achieved. Cells were then pelleted (5 min, 350*g*) and re-suspended in serum-free medium. The culture was collected when cell viability dropped below 30% (after 5–7 days). Monoclonal antibody 83-14 from collected supernatant was captured using a ProSep-vA column (Millipore). Monoclonal antibody 83-14 was proteolysed with dithiothreitol-activated papain (Sigma-Aldrich) at 37 °C. Proteolysis was stopped by adding IAA and the reaction mixture containing Fab 83-14 was then passed down a Superdex 200 26/60 column. Final purification of Fab 83-14 was by cation-exchange chromatography on a MonoS column (GE Healthcare).

Purification of the IR593.αCT + Fab 83-14 + insulin complex. The IR593.αCT protein was combined with an excess of bovine insulin and Fab 83-14 and then re-subjected to SEC. The SEC profiles (Supplementary Fig. 8i) showed a major peak with a molecular mass corresponding to that of a dimer of Fab-complexed IR593.αCT homodimers and a shoulder of molecular weight corresponding to that of an Fab-complexed IR593.αCT homodimer, as well as peaks corresponding to excess Fab 83-14 and insulin. Fractions corresponding to the higher molecular mass component of the major peak were pooled and then concentrated in a 10 mM HEPES-NaOH (pH 7.5) + 0.02% NaN_3 buffer for crystallization.

Production of IR468-containing media. A DNA fragment encoding the signal peptide plus residues 1–468 of human insulin receptor was inserted into the HindIII–XbaI sites of the vector pcDNA3.1zeo(+) (Life Technologies) and then transiently transfected into HEK293 cells. Cells were grown in Dulbecco's modified Eagle's medium (Life Technologies) supplemented with 10% FCS for 72 h after transfection; media containing IR468 were then collected and stored at −20 °C until used.

Crystallization and diffraction data collection of complexes A to D. Complexes A, B, C and D are as defined in Table 1. Complexes A, B and C were prepared for crystallization by combining the Fab 83-7 bound IR310.T (final concentration 3.5 mg ml^{-1}) with a 1.5× molar ratio of αCT peptide (Genscript) and 3× molar ratio of human insulin (Sigma-Aldrich) or [D-Pro^{B26}]-DTI-NH_2 in 10 mM HEPES-NaOH (pH 7.5) + 0.02% sodium azide (NaN_3). [D-Pro^{B26}]-DTI-NH_2 was prepared as described previously[37]. A single crystallization condition for complex A was detected using a sparse-matrix 792-condition screen (CSIRO Collaborative Crystallization Centre) and optimized manually to 0.9–1.1 M tri-sodium citrate, 0.1 M imidazole-HCl (pH 8.0) + 0.02% NaN_3 for all complexes A, B and C. An initial crystallization condition for complex D at 2.0 mg ml^{-1} was detected in a similar 792-condition screen and optimized manually to 9% PEG3350, 200 mM proline, 100 mM HEPES-NaOH (pH 7.5). For data collection, the complex A and B crystals were cryo-protected by transfer to paraffin oil HR403 (Hampton Research), the complex C crystal by progressive increase of the precipitant concentration to saturated tri-sodium citrate, and complex D crystal by addition of 20% glycerol to the crystallant solution. All crystals were cryo-cooled by plunging into liquid nitrogen and maintained at ~100 K during data collection. Data set 1 for the complex A crystal and the data set for the complex D crystal were collected at beamline MX2 at the Australian Synchrotron[38] ($\lambda = 0.95369$ Å). The remaining data sets were collected at beamline I24 at the

Diamond Light Source ($\lambda = 0.96860$ Å). Diffraction data were processed with XDS[39] and CCP4[40]. Resolution limits were assessed on the basis of the significance of $CC_{1/2}$ at the $P = 0.001$ level[19]. Data processing statistics are provided in Table 1.

Structure solution and refinement of complex A. PHASER[41] located within the asymmetric unit a single copy each of the L1–CR module and the variable module of 83-7 (using data set 1). The 83-7 constant module was not located and ultimately presumed disordered. TLS parameters and individual restrained isotropic B-factors were then refined using autoBUSTER[42], followed by atomic coordinate-only refinement, yielding $R^{xpct}{}_{work}/R^{xpct}{}_{free} = 0.368/0.363$ (where R^{xpct} is the crystallographic R-factor reported by autoBUSTER, computed using the expectation value of F_{calc} instead of the value itself)[43]. The difference density maps contained four helix-like features encompassing the first, third, fourth, fifth and seventh highest peaks in the map (the second highest peak being glycan conjugated to Asn 111; Supplementary Fig. 10a). These features lay adjacent to the L1–β_2 surface, and PHASER then positioned the three insulin helices as a rigid entity onto three of these (TFZ = 11.4). Correctness of this solution was verified by exhaustive six-dimensional real-space search using ESSENS[44] (Supplementary Fig. 10b). The structure was then refined further, yielding $R^{xpct}{}_{work}/R^{xpct}{}_{free} = 0.320/0.339$. The remaining helix-like feature in the above difference map (the first and fifth highest peaks) revealed clear side-chain protrusions upon negative B-factor map sharpening[21], spaced consistent with an underlying α-helix (Supplementary Fig. 10c). We concluded that it arose from the αCT peptide as there were no further sources of peptide in the crystallization sample. A ten-residue poly-alanine α-helix was then docked into this feature using ESSENS; the best fit was adequately discriminated from lower-scoring fits and visual inspection confirmed the direction of the helix as judged by protruding side-chain density (Supplementary Fig. 10c). A procedure was then devised to assign, in an objective fashion, the sequence register to this 10-residue segment, based on compatibility of individual residue side chains both with sharpened difference electron density and with their surrounding protein environment. (1) Fit to difference density used a method designed not only to assess the fit of atoms within the density, but also to penalize the existence of volumes of positive difference density into which no atoms had been placed. Briefly, voxels corresponding to the αCT feature were excised from the B-factor-sharpened[21] map using CHIMERA[45] (cutoff level = 0.16 e^{-} Å^{-3}) and placed within a rectangular grid volume large enough to allow a >8 Å buffer zone around the feature. All grid points outside of the feature were set to 0 e^{-} Å^{-3}, with the resultant map being termed M_{obs}. Coordinates for residues 1, 2 and 10 of the fitted poly-alanine helix were then deleted from the model, as the density associated with them displayed poorer α-helical geometry. Ten alternatives, corresponding to all possible alignments of sub-sequences of αCT(704–719) with the heptameric poly-alanine structure, were then assessed. Rotamers for each residue within each of these ten models were then assigned manually using COOT[46] based on visual inspection of the fit of the trial rotamer to the density at the corresponding site in M_{obs}. The ten individual models were then 'real-space-refined' within COOT to achieve optimal fit to M_{obs}, maintaining tight helical restraints. An electron density map M_{calc}, on the same grid as M_{obs}, was then generated (using SFALL within CCP4) for each heptamer model in isolation, with the B-factors of all main-chain atoms being set to 10 Å^2 and of all side-chain atoms set to 20 Å^2 to allow for subsequent comparison with a sharpened map. All grid points within M_{calc} with $\rho < 0.45$ e^{-} Å^{-3} were set to 0 e^{-} Å^{-3}. Correlation coefficients $CC = (<xy> - <x><y>)/\sqrt{[(<x^2> - <x>^2)(<y^2> - <y>^2)]}$ between M_{obs} and each M_{calc} were then calculated using MAPMAN[47]. Each CC was then 'normalized' by dividing its value by the CC value calculated for the underlying poly-alanine heptamer, and the quotient termed the trial sequence's 'density score'. The heptamer with trial register 705–711 was seen to have the highest density score (Supplementary Table 4). (2) Compatibility of residue side-chain environment within the L1–αCT–insulin interface was assessed as follows. For each of the ten trial heptamer (real-space refined) models described above, an 'environment score' was generated using VERIFY3D[48] to assess compatibility with the surrounding L1–β_2 and insulin surfaces. Trial register 705–711 again scored highest (Supplementary Table 4). A 'combined score' was then computed as the product (density score) $\times$ (environment score) to assist assessment of lower-ranked trial registers (Supplementary Table 4). The next highest combined score was for trial register 709–715, which is related to trial register 705–711 by a one-turn translation along the helix, effectively maintaining hydrophobic-to-hydrophobic docking with the L1 surface. However, register 709–715 was judged to be most unlikely, as it would bring Pro 716 and Pro 718 into the remaining C-terminal region of the helix. We concluded that register assignment 705–711 was correct, given that it ranked highest on all criteria. The assignment aligned with that assessed directly from the EDDMs for complex D, wherein the αCT segment is directly attached to the C-terminus of FnIII-1 and the ambiguity of assignment strictly limited. Nine sugar residues were also included at N-linked sites 16, 25, 111, 225 and 255[16,49]. Final refinement was against data set 2, using local structural similarity restraints (LSSRs)[22] against Protein Data Bank (PDB) structures 2HR7-A (L1-CR), 1IL1-B (83-7 light chain) and 1FNS-H (83-7 heavy chain), yielding $R^{xpct}{}_{work}/R^{xpct}{}_{free} = 0.264/0.292$ (see Supplementary Table 1). Ramachandran statistics are: favoured region 91%, acceptable region 5% and disallowed region 4%.

Structure solution and refinement of complexes B and C. These structures were solved independently. MOLREP[50] was used to locate the L1–CR/83-7 fragment and the structure then refined with REFMAC5[23]. [D-Pro^{B26}]-DTI-NH_2 (PDB entry 2WRW) was then located using MOLREP, its position being effectively identical to that in the native insulin complex (complex A). The αCT segment from the wild-type insulin complex was fitted in the final electron density maps using COOT. Final refinement gave $R_{work}/R_{free} = 0.289/0.349$ and $R_{work}/R_{free} = 0.289/0.335$, respectively (Supplementary Table 1). Ramachandran statistics for both complexes are: favoured region 89%, acceptable region 7% and disallowed region 4%.

Structure solution and refinement of complex D. Molecular replacement with PHASER located sequentially two copies each of a module consisting of Fab 83-14 in complex with the FnIII-1 domain, an L1–CR module and an L2 domain, using search models derived from PDB entry 3LOH[12]. The assembly was subject to rigid-body and TLS refinement using autoBUSTER, followed by individual restrained isotropic B-factor and coordinate refinement, subject both to LSSRs to PDB entries 2HR7-A (L1-CR-L2), 3LOH-E (FnIII-1), 1PLG-H (83-14 heavy chain) and 2VXU-M (83-14 light chain) and to non-crystallographic symmetry (NCS) restraints, yielding $R^{xpct}{}_{work}/R^{xpct}{}_{free} = 0.308/0.299$. Difference electron density analysis revealed: (1) two helix-like features of electron density, each connected to a respective insulin receptor residue Thr 593 and lying on the L1–β_2 surface of an adjacent monomer, thus identifiable as the αCT segment of IR593.αCT; and (2) two clusters of three helix-like features, each cluster lying adjacent to a respective one of the above αCT segments and putatively identifiable as insulin. Insulin was then docked into each of the above putative insulin–αCT clusters of features by means of an exhaustive six-dimensional ESSENS search. The highest scoring solution in each instance overlaid insulin accurately onto the respective three-helix cluster (Supplementary Fig. 10d) and was well discriminated from lower scoring solutions; the pose of insulin and αCT with respect to L1–β_2 was effectively identical at the two NCS related sites. The two insulins were then included in the model followed by further refinement. COOT was then used to position poly-alanine α-helices into the respective density features corresponding to the αCT segment (sharpened by the application of a negative B-factor), followed by mutation of the sequence to that of αCT residues 704–715. Assignment of the αCT register presented no more than a single residue prima facie ambiguity (depending on the way in which residues at the FnIII-1–αCT peptide junction were built). One alternative was strongly favoured, aligning independently with that determined for complexes A, B and C above. Residues B5–B7 of each insulin could be built into electron density following a path similar to that seen in certain high-affinity insulin analogues[18]. Further crystallographic refinement was then undertaken with autoBUSTER, yielding final $R^{xpct}{}_{work}/R^{xpct}{}_{free} = 0.244/0.260$ (Supplementary Table 1). Ramachandran statistics are favoured region 88%, acceptable region 7% and disallowed region 5%.

Holoreceptor photo-crosslinking. Production and photo-crosslinking of derivatized holoreceptors were performed as described previously[25].

Photo-crosslinking of Pap-derivatized insulins to IR468. Photo-activatable insulin derivatives Pap^{A3}-Ins, Pap^{A8}-DKP-Ins, Pap^{B16}-DKP-Ins, Pap^{B24}-DKP-Ins, [Pap-^{B25}]-DP-Ins were synthesized as described[28,51] and iodinated using the chloramine-T oxidation method. αCT(703–719)-Myc peptide was synthesized in-house. ^{125}I-photo-activatable insulin analogues were incubated overnight at 4 °C with IR468 and 10^{-5} M αCT(703–719)-Myc in 100 mM HEPES (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$, 0.05% bovine serum albumin, 0.025% Triton X-100. One aliquot was exposed to short-wavelength ultraviolet light (Mineralight UVG-54) for 30 s at 1 cm and a second aliquot treated with 0.2 mM disuccinimidyl suberate (DSS; Pierce). Samples were diluted with Laemmli sample buffer, resolved on 10% Tricine SDS-PAGE and auto-radiographed.

Isothermal titration calorimetry (ITC). ITC analysis of insulin binding to IR310.T and its Fab 83-7 complex followed protocols identical to those described previously[52]. The affinity of zinc-free human insulin for IR310.T in the presence of a 10× molar ratio of αCT(704–719) was determined to be $K_d = 30$ nM (Supplementary Fig. 8e) and the affinity of zinc-free human insulin for IR310.T–Fab 83-7 in the presence of a 10× molar ratio of αCT(704–719) to be $K_d = 48$ nM (Supplementary Fig. 8f). ITC analysis of mutant αCT peptides binding to IR485 and of insulin binding to IR485 in complex with mutant αCT peptide were also performed as described previously[52], with samples prepared in TBSA (25 mM Tris-HCl, 137 mM sodium chloride, 2.7 mM potassium chloride, pH 8.0 + 0.02% NaN_3).

31. Hoogenboom, H. R. *et al.* Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* **19,** 4133–4137 (1991).

32. Stanley, P. Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol. Cell. Biol.* **9,** 377–383 (1989).
33. Soos, M. A. *et al.* Monoclonal antibodies reacting with multiple epitopes on the human insulin receptor. *Biochem. J.* **235,** 199–208 (1986).
34. Evan, G. I., Lewis, G. K., Ramsay, G. & Bishop, J. M. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.* **5,** 3610–3616 (1985).
35. Denley, A. *et al.* Structural determinants for high-affinity binding of insulin-like growth factor II to insulin receptor (IR)-A, the exon 11 minus isoform of the IR. *Mol. Endocrinol.* **18,** 2502–2512 (2004).
36. Markussen, J., Halstrom, J., Wiberg, F. C. & Schäffer, L. Immobilized insulin for high capacity affinity chromatography of insulin receptors. *J. Biol. Chem.* **266,** 18814–18818 (1991).
37. Žáková, L. *et al.* Insulin analogues with modifications at position B26. Divergence of binding affinity and biological activity. *Biochemistry* **47,** 5858–5868 (2008).
38. McPhillips, T. M. *et al.* Blu-ice and the distributed control system: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **9,** 401–406 (2002).
39. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66,** 133–144 (2010).
40. Collaborative Computing Project, No The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
41. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63,** 32–41 (2007).
42. Bricogne, G. *et al.* BUSTER version 2.10 (Global Phasing Ltd, 2011).
43. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60,** 2210–2221 (2004).
44. Kleywegt, G. J. & Jones, T. A. Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallogr. D* **53,** 179–185 (1997).
45. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comp. Chem.* **25,** 1605–1612 (2004).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
47. Jones, T. A. & Thirup, S. Using known substructures in protein model building and crystallography. *EMBO J.* **5,** 819–822 (1986).
48. Lüthy, R., Bowie, J. U. & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **356,** 83–85 (1992).
49. Sparrow, L. G. *et al.* N-linked glycans of the human insulin receptor and their distribution over the crystal structure. *Proteins Struct. Funct. Bioinform.* **71,** 426–439 (2008).
50. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66,** 22–25 (2010).
51. Hua, Q. X. *et al.* Enhancing the activity of a protein by stereospecific unfolding. The conformational life cycle of insulin and its evolutionary origins. *J. Biol. Chem.* **248,** 14586–14596 (2009).
52. Menting, J. G., Ward, C. W., Margetts, M. B. & Lawrence, M. C. A thermodynamic study of ligand binding to the first three domains of the human insulin receptor: relationship between the receptor α-chain C-terminal peptide and the site 1 insulin mimetic peptides. *Biochemistry* **48,** 5492–5500 (2009).

doi:10.1038/nature11786

Glaciations in response to climate variations preconditioned by evolving topography

Vivi Kathrine Pedersen[1] & David Lundbek Egholm[2]

Landscapes modified by glacial erosion show a distinct distribution of surface area with elevation[1–3] (hypsometry). In particular, the height of these regions is influenced by climatic gradients controlling the altitude where glacial and periglacial processes are the most active, and as a result, surface area is focused just below the snowline altitude[1–9]. Yet the effect of this distinct glacial hypsometric signature on glacial extent and therefore on continued glacial erosion has not previously been examined. Here we show how this topographic configuration influences the climatic sensitivity of Alpine glaciers, and how the development of a glacial hypsometric distribution influences the intensity of glaciations on timescales of more than a few glacial cycles. We find that the relationship between variations in climate and the resulting variation in areal extent of glaciation changes drastically with the degree of glacial modification in the landscape. First, in landscapes with novel glaciations, a nearly linear relationship between climate and glacial area exists. Second, in previously glaciated landscapes with extensive area at a similar elevation, highly nonlinear and rapid glacial expansions occur with minimal climate forcing, once the snowline reaches the hypsometric maximum. Our results also show that erosion associated with glaciations before the mid-Pleistocene transition at around 950,000 years ago probably preconditioned the landscape—producing glacial landforms and hypsometric maxima—such that ongoing cooling led to a significant change in glacial extent and erosion, resulting in more extensive glaciations and valley deepening in the late Pleistocene epoch. We thus provide a mechanism that explains previous observations from exposure dating[10] and low-temperature thermochronology[11] in the European Alps, and suggest that there is a strong topographic control on the most recent Quaternary period glaciations.

It has long been recognized that topographic feedbacks between glacial erosion and glacial mass balance influence glaciations[7,12–20]. For example, previous numerical studies have shown how glacial erosion can reduce the extent of subsequent glaciations under constant climate conditions by lowering of topography[13,18]. However, the effect of a distinct glacial landscape hypsometry on glacial extent through several glacial cycles has not been considered explicitly, and limited work only has focused on transient effects in glacial extent resulting from glacial erosion processes modifying the topography[13,16–20].

Here we examine the response of mountain-range glaciations to the present topographic distribution for a varying climate using a numerical surface process model, including an ice-sheet model suitable for rugged mountain topography[21–23] (see Methods Summary and Supplementary Information). Furthermore, in an additional experiment, we impose a climate forcing comparable with Quaternary climate records[24], to investigate possible feedbacks between climate, topography, glacial extent and glacial erosion that may have prevailed throughout the last two million years (Myr) of Earth's history.

The hypsometry of a landscape influences the mass balance of glaciers, because it dictates how much surface area is available for snow and ice accumulation when paired with a temperature–elevation distribution. Therefore, for a given variation in snowline altitude, the change in glacier accumulation area is a nonlinear function of the topographic configuration. This nonlinearity becomes increasingly pronounced through the development of the distinct hypsometric maximum found worldwide below the present snowline altitude for glacially modified areas[2,3] (see Supplementary Figs 1–5).

To illustrate this, we use our numerical approach for simulating glaciation of a catchment from Sierra Nevada, Spain, where limited glacial activity has occurred throughout the Quaternary[25], and of a catchment from the Bitterroot Range, USA, which has been significantly modified by glaciers during the Quaternary[26] (see Methods Summary, Supplementary Information, and Supplementary Videos 1 and 2). Modelled glaciers develop as a result of steady cooling over a period of 50 thousand years (kyr) followed by a 50-kyr steady temperature increase. The snowline altitude, which in our model is a linear function of surface temperature, is for both catchments lowered from 2,800 m down to 1,900 m (Fig. 1a, b). This snowline span of 900 m corresponds roughly to the difference between present-day snowline altitudes and suggested Last Glacial Maximum (LGM) snowline altitudes[27]. Importantly, although it is representative of the LGM snowline variation in the Bitterroot Range, the snowline interval is not meant to represent the actual LGM snowline lowering for Sierra Nevada, where the real snowline altitude reached only the highest parts of the mountain range during the Quaternary[25]. Instead, the experiment is designed to highlight the isolated effect of catchment hypsometry on glacial extent by keeping everything else equal. For the same reason, no tectonic uplift or erosion is introduced.

For the fluvial catchment in Sierra Nevada, the ice volume increases with a slightly accelerating rate during snowline lowering as the amount of surface area increases downwards (Fig. 1c, d, Supplementary Video 1). However, for the glacially modified catchment in Bitterroot Range, the relationship between snowline lowering and glacial extent is highly nonlinear. The ice volume increases drastically when the snowline altitude reaches the hypsometric maximum (Fig. 1c, Supplementary Video 2). However, as the snowline altitude subsequently falls below the hypsometric maximum, the rate of change in ice volume decreases to values lower than for the Sierra Nevada catchment (Fig. 1d). So, for a similar climatic forcing, the resulting ice volume (that is, areal extent) is very sensitive to the topographic distribution in the two catchments. The nonlinearity between climate change and glacial extent found for glacially modified landscapes is activated when the snowline altitude varies in proximity to the distinct glacial hypsometric maximum, which would be the case for a climate cooling from the present-day level.

We then examine the transient signal in glacial extent and glacial erosion for a climate forcing similar to what is suggested to have prevailed throughout the last 2 Myr of the Quaternary[24] by introducing glaciers and glacial erosion in a simulated fluvial steady-state landscape (Fig. 2a, f, and Supplementary Videos 3 and 4). We reproduce the first-order patterns of the Quaternary climate variations by introducing, first, a phase of symmetric constant-magnitude 40-kyr climate cycles (phase 1), followed by a series of asymmetric 100-kyr climate cycles with increasing amplitude (phase 2), leading to an overall decrease in

[1]Department of Earth Science, University of Bergen, Allégaten 41, 5007 Bergen, Norway. [2]Department of Geoscience, Aarhus University, Høegh-Guldbergs Gade 2, 8000 Aarhus, Denmark.

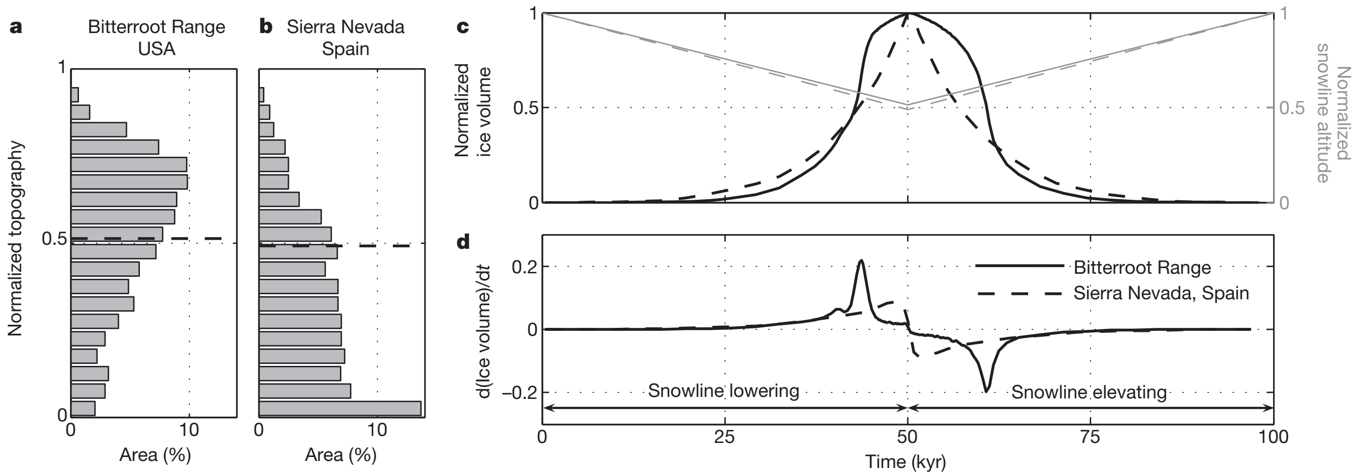


Figure 1 | Ice volume as a function of climate change and hypsometry. **a**, **b**, Hypsometric distributions for the glacially modified Bear Creek catchment in the Bitterroot Range, Montana, USA (46.4° N, 114.4° W), and the fluvial Aldeire catchment from Sierra Nevada, Spain (37.1° N, 3.1° W). The topography has been normalized to the maximum height above the local base level. Black dashed lines represent the minimum level of the snowline altitude at 50 kyr. For a map view see Supplementary Figs 11 and 12. **c**, Temporal evolution in ice volume for the catchments in Bitterroot Range (solid black line) and in Sierra Nevada (dashed black line) for a changing snowline altitude (solid grey line represents the Bitterroot Range and dashed grey line the Sierra Nevada). **d**, Temporal evolution of the rate of ice volume change for the two landscapes. Animations (Supplementary Video 1 and 2) are available online, showing simulated glaciations in the two catchments.

mean temperature, before a final phase of asymmetric constant-magnitude 100-kyr climate cycles (phase 3, Fig. 2a). To expose the effect of glacial erosion on the areal extent of glaciations, our experiment includes only glacial erosion processes and the resulting flexural isostatic responses of both loading and unloading due to ice and erosion. The change in ice volume (that is, glacial extent) is shown relative to a reference simulation without glacial erosion (Fig. 2b–d).

During the first phase of constant-magnitude climate cycles (Fig. 2a), landforms that are typically associated with a glacial origin emerge. The numerical model includes higher-order ice physics, which facilitates a detailed high-resolution simulation of landscape evolution on scales small enough to resolve U-shaped valleys, hanging valleys, cirques and steep headwalls (Fig. 2h, Supplementary Fig. 14). Furthermore, a hypsometric maximum develops just below the minimum imposed snowline altitude, similar to global observations (Fig. 2e, Supplementary Fig. 1). Concurrently with the development of these topographic features, the ice volume, and therefore glacial erosion, decreases throughout the occurring glaciations as a consequence of glacial erosion, irrespective of climate being constant for each cycle (Figs 2b and 3a). Erosion is, to a high degree, focused in the steep regions above the lowest snowline altitude, where glacial conditions prevail even during the warmest periods of each climate cycle (Fig. 3b). Glacial erosion below the lowest snowline altitude is limited to major valleys, and glaciers reach far below this level only during the coldest periods of the first climate cycles. Thus, glacial erosion lowers mainly the topography at high elevations, reducing the glacier accumulation space and initiating a negative feedback between glacial erosion and glacial extent[28].

During the second phase (Fig. 2a) mimicking the mid-Pleistocene transition from 40-kyr to 100-kyr climate cycles, the ice volume increases as a consequence of cooling. More interestingly, the change in ice volume is greater for the glacially modified landscape than for the reference model, and the model with glacial erosion reaches even higher ice volumes than the model without glacial erosion (Fig. 2b, d). In other words, the glaciated cover of the glacially eroded landscape is more sensitive to climatic variations because a lot of surface area is concentrated within a narrow elevation span in the same way as is seen for the Bitterroot catchment compared to the fluvial Sierra Nevada (Fig. 1). Another outcome of focusing surface area within a narrow elevation interval is that the glaciations following glacial erosion initiate later than for the reference model without glacial erosion, suggesting shorter but more vigorous glaciations in response to the same climatic forcing (Fig. 2b, d). During the constant-magnitude climate cycles in the third phase, the negative feedback between glacial erosion and glacial extent returns, although it is not as pronounced as in phase 1. This is because erosion in phase 1 has already modified the longitudinal and transverse shapes of valleys at high elevations, and these landforms are now associated with smaller subglacial erosion rates than their fluvial counterparts (Supplementary Figs 7 and 15).

As a consequence of the strong negative feedback between glacial erosion, glacial mass balance, and ice extent in phase 1, only limited glacial erosion occurs in the main valley during the constant-magnitude glacial cycles of this phase (Fig. 3a, b). The negative feedback is reflected by the decreasing glacial erosion rate as well as the shorter glacial cycles (Fig. 3a). However, as general cooling is introduced in phase 2, the negative feedback is broken because snowline lowering reincorporates the previously lowered topography into the area available for ice accumulation. This leads to a significant jump in glacial extent across the transition from the ice-poor conditions at the end of phase 1 to the large glaciations in phases 2 and 3 (Fig. 2c), enhancing the potential for glacial erosion in the main valleys (Fig. 3c). The model experiments thus reveal a greater effect of cooling across the mid-Pleistocene transition when the landscape has been prepared by earlier glaciations (Fig. 2c and Supplementary Fig. 28). That the model including glacial erosion has even more ice in phases 2 and 3 than the model without erosion is caused by flexural isostatic rock uplift, adding surface area to the hypsometric maximum from below, in combination with glacially flattened longitudinal valley profiles (Supplementary Figs 15, 17 and 19), which require thicker ice to maintain ice flux from the accumulation zone to the ablation area.

The results presented here thus suggest that topographic modifications occurring throughout the Quaternary may have led to a change in glaciations for the following cycles, by promoting an increasingly nonlinear relation between climate change and the area available for snow and ice accumulation. In particular, as the climate cooled generally throughout the Quaternary, the more extensive glaciations found in the final part of the Quaternary and the contrast in glaciations and glacial erosion across the mid-Pleistocene transition may not only be the direct results of colder temperatures, but also the responses to the developing glacial topographic signature in combination with snowline lowering. The significant increase in simulated glacial extent and glacial erosion across the mid-Pleistocene transition (Figs 2c

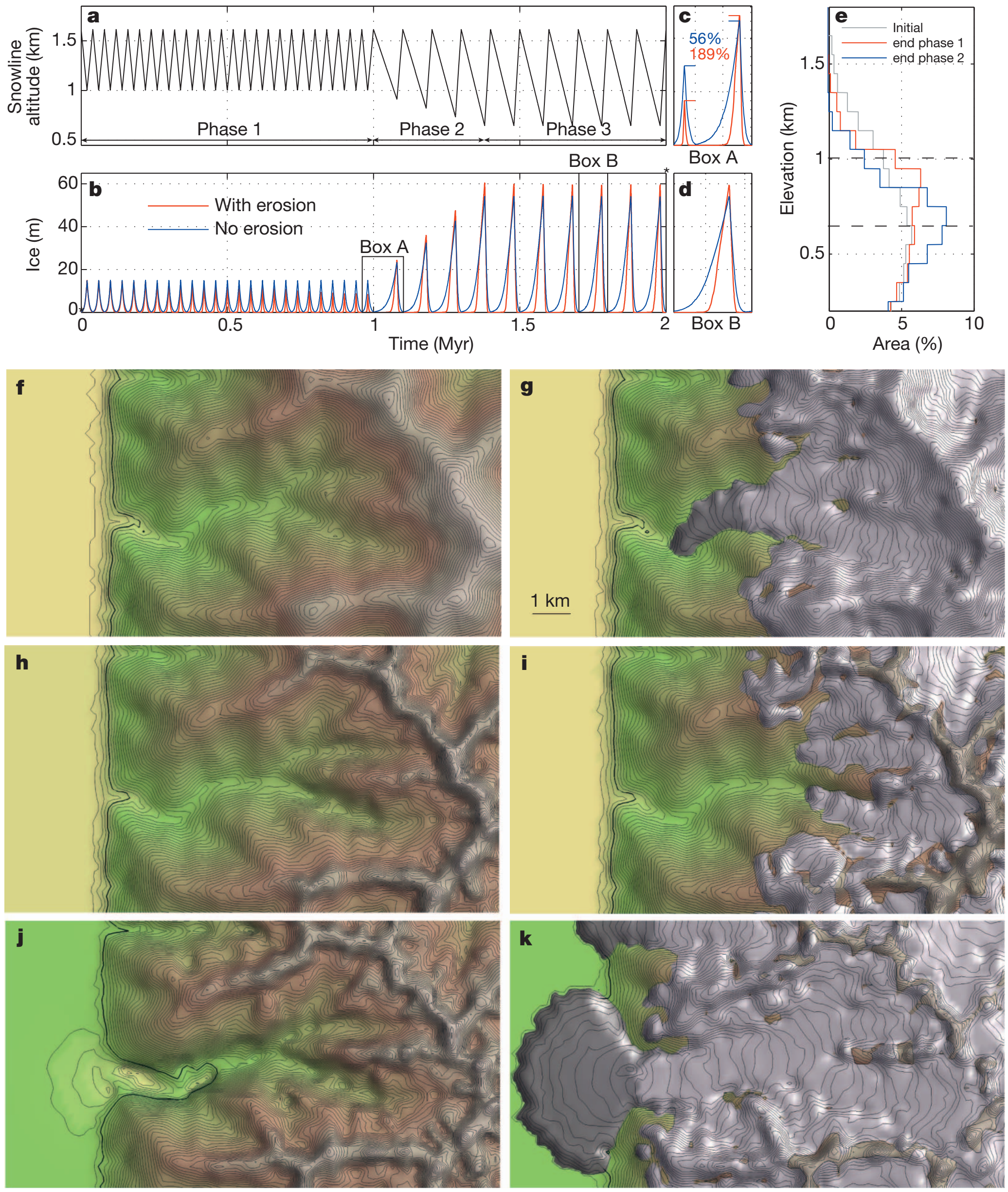


Figure 2 | Transient changes in ice volume for a Quaternary-like climate forcing. **a**, Temporal evolution in snowline altitude. **b**, Temporal evolution in spatially averaged ice volume for models with (red) and without (blue) glacial erosion. For our choice of glacial erosion rate, see experiment S1 in the Supplementary Information. **c**, Enlarged view of the change in ice volume on the transition from phase 1 to phase 2 for the models with (red) and without (blue) glacial erosion, respectively (Box A). **d**, Enlarged view of the glacial cycle in Box B in **b**). **e**, Hypsometric distribution for the initial fluvial landscape (grey), the landscape after phase 1 (red), and the final landscape after phase 3 (blue). The dashed black lines represent the minimum snowline altitude in phases 1 and 3, respectively. **f**, Initial fluvial steady-state landscape. Contour spacing is 25 m. The 100-m contour has been highlighted for reference. **g**, Glacial extent in the first glacial cycle at 20 kyr. **h**, Landscape configuration after phase 1. **i**, Glacial extent in the last glacial cycle of phase 1 at 980 kyr. **j**, Final landscape after phase 3. Note the uplifted foreland caused by flexural isostatic effects. **k**, Glacial extent in the final glacial cycle at 1,980 kyr. Details of modelling parameters, full size figures, and animations are available in the Supplementary Information.

and 3, and Supplementary Fig. 28) can explain results found using exposure dating[10] and low-temperature thermochronology[11] in the European Alps, suggesting an increase in relief around a million years ago owing to extensive valley deepening and only minor erosion at higher elevations in association with extensive glaciations following the mid-Pleistocene transition.

In another new study comparing offshore sediment volumes with onshore estimates of fjord erosion in Norway[29], it has been suggested that extensive erosion occurred both in fjords and at higher elevations in the Quaternary. This study may therefore record both an initial phase of high-elevation erosion and a subsequent extensive deepening of valleys. By suggesting two different timings of main cirque formation and main valley deepening, our results therefore reconcile both the hypothesis of bimodal Quaternary glacial erosion[17,29] and the studies showing an increased relief across the mid-Pleistocene transition[10,11].

Climate change is the main driver for glaciations. However, as we have demonstrated here, the topographic distribution that emerges as a result of glacial erosion processes in alpine settings influences the relation between climate and glacial extent in a highly nonlinear manner by changing the area available for snow and ice accumulation. The fact that areas modified recently by glacial erosion processes generally have a distinct hypsometric distribution, with a maximum just below the present day snowline altitude (Supplementary Fig. 1),

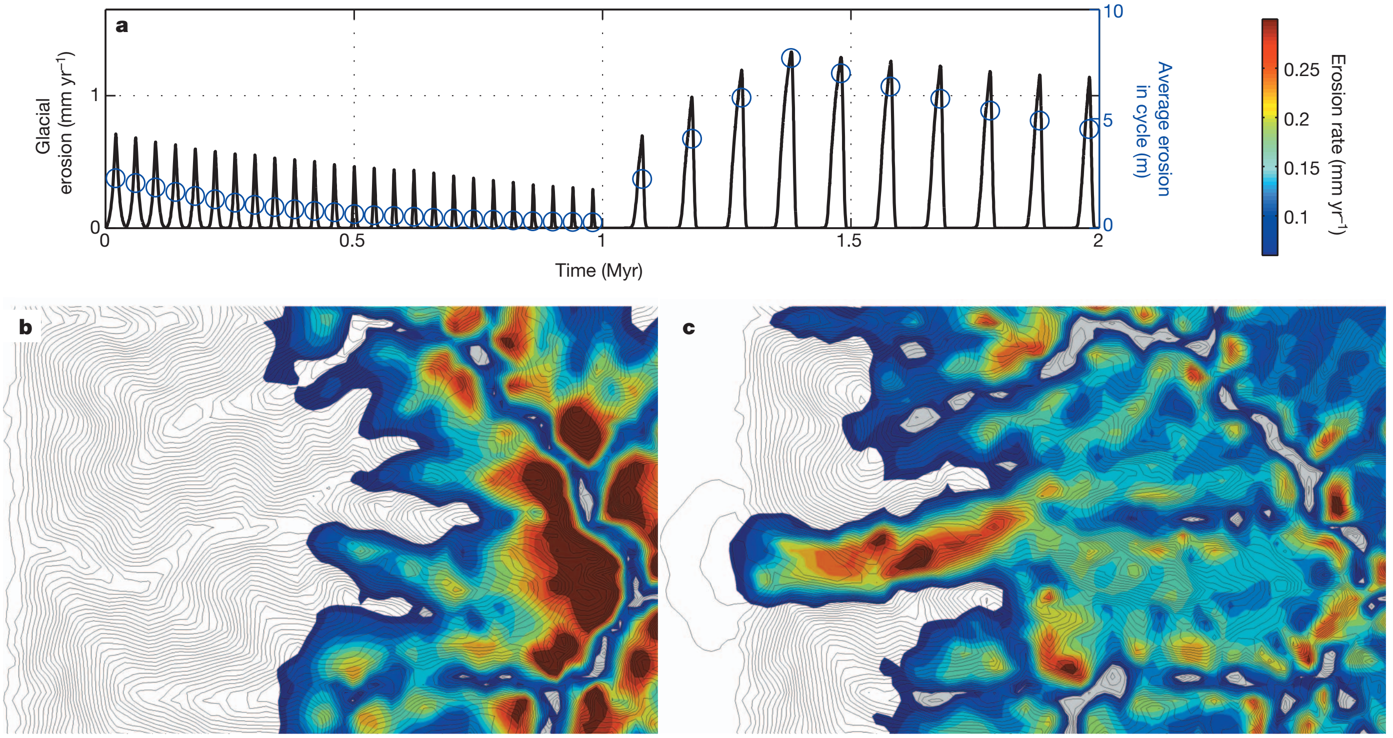


Figure 3 | Variations in glacial erosion for a Quaternary-like climate. **a**, Temporal variation in spatially averaged glacial erosion rate (black line) and average glacial erosion for each glacial cycle (blue circles). **b**, Average glacial erosion rate in phase 1. **c**, Average glacial erosion rate in phases 2 and 3. Spatial glacial erosion patterns are available for representative cycles in Supplementary Figs 24–27. For comparison with glacial erosion rates in the initial fluvial landscape see Supplementary Fig. 28. Erosion histories for representative locations are presented in Supplementary Fig. 29.

suggest that these nonlinear effects may have affected most regions occupied by alpine-type glaciations throughout the Quaternary, and that recent fluctuations in areal extent of glaciations may have been significantly affected by a constantly developing topographic control on glacier mass balance.

METHODS SUMMARY

The numerical surface process model (SPM) we used couples the flow of water, ice and sediment. The flow of ice is handled using a higher-order ice-sheet model, taking into account effects from steep gradients in the bed and ice surfaces, and internal stresses in the ice[21,22] (integrated Second Order Shallow Ice Approximation; iSOSIA). These effects are of importance when modelling ice flow in alpine settings, and their implementation facilitates detailed predictions of long-term landscape evolution for direct comparison with present-day landforms. On the basis of the flow components, fluvial erosion, glacial erosion (abrasion/quarrying), and mass-wasting processes can be estimated using a set of erosion laws[23]. Tectonic uplift and flexural isostatic effects related to loading/unloading of ice and erosion are included. The combined model set-up has been described elsewhere[21–23], including testing of the iSOSIA approach against other higher-order ice models as well as full-Stokes solutions for ice flow[21], however, a summary of the set-up and chosen parameter values can be found in the Supplementary Information.

For the glaciation experiments of real catchments from Sierra Nevada (Spain) and the Bitterroot Range (USA), extracts from the SRTM3[30] data set were interpolated onto a SPM grid with similar resolution. The experiment illustrating Quaternary evolution (Figs 2 and 3) initiates from a synthetic fluvial steady state, where rock column uplift is balanced by erosion caused by fluvial and hillslope processes. This state is reached by uplifting the model domain (20 km × 20 km) at a constant uniform rate, while fixing the boundary nodes at sea level. The actual experiment excludes tectonic uplift and erosion by fluvial and hillslope processes, except that a local threshold criterion, representing local short-term hillslope processes, prevents the emergence of unrealistic steep surface slopes. Lowland is added to the edge of the model domain in order for ice to extend from the topography.

Received 18 June; accepted 8 November 2012.

1. Brocklehurst, S. H. & Whipple, K. X. Hypsometry of glaciated landscapes. *Earth Surf. Process. Landf.* **29,** 907–926 (2004).
2. Egholm, D. L., Nielsen, S. B., Pedersen, V. K. & Lesemann, J. E. Glacial effects limiting mountain height. *Nature* **460,** 884–887 (2009).
3. Pedersen, V. K., Egholm, D. L. & Nielsen, S. B. Alpine glacial topography and the rate of rock column uplift: a global perspective. *Geomorphology* **122,** 129–139 (2010).
4. Brozović, N., Burbank, D. & Meigs, A. Climatic limits on landscape development in the northwestern Himalaya. *Science* **276,** 571–574 (1997).
5. Mitchell, S. G. & Montgomery, D. R. Influence of a glacial buzzsaw on the height and morphology of the Cascade Range in central Washington State, USA. *Quat. Res.* **65,** 96–107 (2006).
6. Montgomery, D. R., Balco, G. & Willett, S. D. Climate, tectonics, and the morphology of the Andes. *Geology* **29,** 579–582 (2001).
7. Whipple, K. X., Kirby, E. & Brocklehurst, S. H. Geomorphic limits to climate-induced increases in topographic relief. *Nature* **401,** 39–43 (1999).
8. Hales, T. C. & Roering, J. J. A frost "buzzsaw" mechanism for erosion of the eastern Southern Alps, New Zealand. *Geomorphology* **107,** 241–253 (2009).
9. Delunel, R., van der Beek, P. A., Carcaillet, J., Bourlès, D. L. & Valla, P. G. Frost-cracking control on catchment denudation rates: insights from *in situ* produced ^{10}Be concentrations in stream sediments (Ecrins–Pelvoux massif, French Western Alps). *Earth Planet. Sci. Lett.* **293,** 72–83 (2010).
10. Haeuselmann, P., Granger, D. E., Jeannin, P.-Y. & Lauritzen, S. E. Abrupt glacial valley incision at 0.8 Ma dated from cave deposits in Switzerland. *Geology* **35,** 143–164 (2007).
11. Valla, P., Shuster, D. L. & van der Beek, P. A. Significant increase in relief of the European Alps during mid-Pleistocene glaciations. *Nature Geosci.* **4,** 688–692 (2011).
12. Kerr, A. Topography, climate and ice masses: a review. *Terra Nova* **5,** 332–342 (1993).
13. Oerlemans, J. Numerical experiments on large-scale glacial erosion. *Z. Gletsch. Glazialgeol.* **20,** 107–126 (1984).
14. Foster, D., Brocklehurst, S. H. & Gawthorpe, R. L. Glacial-topographic interactions in the Teton Range, Wyoming. *J. Geophys. Res.* **115,** F01007 (2010).
15. Brocklehurst, S. H. & Whipple, K. X. Glacial erosion and relief production in the eastern Sierra Nevada, California. *Geomorphology* **42,** 1–24 (2002).
16. Braun, J., Zwartz, D. & Tomkin, J. A new surface-processes model combining glacial and fluvial erosion. *Ann. Glaciol.* **28,** 282–290 (1999).
17. Herman, F., Beaud, F., Champagnac, J.-D., Lemieux, J.-M. & Sternai, P. Glacial hydrology and erosion patterns: a mechanism for carving glacial valleys. *Earth Planet. Sci. Lett.* **310,** 498–508 (2011).
18. MacGregor, K., Anderson, R., Anderson, S. & Waddington, E. Numerical simulations of glacial longitudinal profile evolution. *Geology* **28,** 1031–1034 (2000).
19. Tomkin, J. Feedbacks and the oscillation of ice masses. *J. Geophys. Res.* **108** (B10), 2488 (2003).

20. Yanites, B. J. & Ehlers, T. A. Global climate and tectonic controls on the denudation of glaciated mountains. *Earth Planet. Sci. Lett.* **325/326,** 63–75 (2012).
21. Egholm, D. L., Knudsen, M. F., Clark, C. D. & Lesemann, J. E. Modeling the flow of glaciers in steep terrains: the integrated Second-Order Shallow Ice Approximation (iSOSIA). *J. Geophys. Res.* **116,** F02012 (2011).
22. Egholm, D. L., Pedersen, V. K., Knudsen, M. F. & Larsen, N. K. On the importance of higher order ice dynamics for glacial landscape evolution. *Geomorphology* **141/142,** 67–80 (2012).
23. Egholm, D. L., Pedersen, V. K., Knudsen, M. F. & Larsen, N. K. Coupling the flow of ice, water, and sediment in a glacial landscape evolution model. *Geomorphology* **141/142,** 47–66 (2012).
24. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* **20,** PA1003 (2005).
25. Alberti, A. P., Díaz, M. V. & Chao, R. B. in *Quaternary Glaciations—Extent and Chronology. Part I. Europe* (eds Ehlers, J. & Gibbard, P. L.) 389–394 (2004).
26. Weber, W. M. Correlation of Pleistocene glaciation in the Bitterroot Range, Montana, with fluctuations of Glacial Lake Missoula, Montana. *Bur. Mines Geol. Mem.* **42,** 1–44 (1972).
27. Porter, S. C. Some geological implications of average Quaternary glacial conditions. *Quat. Res.* **32,** 245–261 (1989).
28. Kaplan, M. R., Hein, A. S., Hubbard, A. & Lax, S. M. Can glacial erosion limit the extent of glaciation? *Geomorphology* **103,** 172–179 (2009).
29. Steer, P., Huismans, R. S., Valla, P. G., Gac, S. & Herman, F. Bimodal Plio-Quaternary glacial erosion and low-relief surfaces in Scandinavia. *Nature Geosci.* **5,** 635–639 (2012).
30. Farr, T. G. *et al.* The shuttle radar topography mission. *Rev. Geophys.* **45,** RG2004 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements V.K.P. thanks the Danish Council for Independent Research and Inge Lehmanns Fund for funding this research. D.L.E. acknowledges funding from the Danish Council for Independent Research under the Sapere Aude Programme. We thank S. Brocklehurst and P. van der Beek for reviews that improved the manuscript considerably.

Author Contributions V.K.P. and D.L.E. performed the global topographic analysis. D.L.E. developed the numerical modelling scheme used. V.K.P. performed the modelling. Both authors contributed equally to the design of the study and writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.K.P. (vivi.pedersen@geo.uib.no).

doi:10.1038/nature11794

Craniofacial development of hagfishes and the evolution of vertebrates

Yasuhiro Oisi[1,2], Kinya G. Ota[3], Shigehiro Kuraku[4], Satoko Fujimoto[2] & Shigeru Kuratani[2]

Cyclostomes, the living jawless vertebrates including hagfishes and lampreys, represent the most basal lineage of vertebrates. Although the monophyly of cyclostomes has been supported by recent molecular analyses, the phenotypic traits of hagfishes, especially the lack of some vertebrate-defining features and the reported endodermal origin of the adenohypophysis, have been interpreted as hagfishes exhibiting a more ancestral state than those of all other vertebrates. Furthermore, the adult anatomy of hagfishes cannot be compared easily with that of lampreys. Here we describe the craniofacial development of a series of staged hagfish embryos, which shows that their adenohypophysis arises ectodermally, consistent with the molecular phylogenetic data. This finding also allowed us to identify a pan-cyclostome pattern, one not shared by jawed vertebrates. Comparative analyses indicated that many of the hagfish-specific traits can be explained by changes secondarily introduced into the hagfish lineage. We also propose a possibility that the pan-cyclostome pattern may reflect the ancestral programme for the craniofacial development of all living vertebrates.

Recent molecular phylogenetic data revealed that extant vertebrates consist of two monophyletic groups: the living jawless vertebrates (cyclostomes) and jawed vertebrates[1–5] (Fig. 1). Most extant vertebrates belong to the gnathostomes, which are characterized by the possession of a jaw and two nostrils[6] (Fig. 1i, j). In contrast, cyclostomes, comprising only lampreys and hagfishes, lack a biting jaw and possess only a single nostril[6] (Fig. 1a–g). Morphologically, however, several traits in hagfishes have been interpreted as representing more ancestral states than found in all other vertebrates, including lampreys[7–12].

In particular, the craniofacial pattern of hagfishes differs fundamentally from that of lampreys, making it difficult to establish morphological homologies between the two cyclostomes[6–8]. Typically, the nasohypophyseal duct ends blindly in lampreys, whereas that in hagfishes continues into the pharynx, exhibiting a conspicuous anatomical difference[13] (Fig. 1c, g and Supplementary Fig. 1). Another impediment to the comparison is linked to whether the hagfish adenohypophysis is derived from the endoderm, similar to its potential homologue of amphioxus[7–10,14,15], or the ectoderm[16], like that of

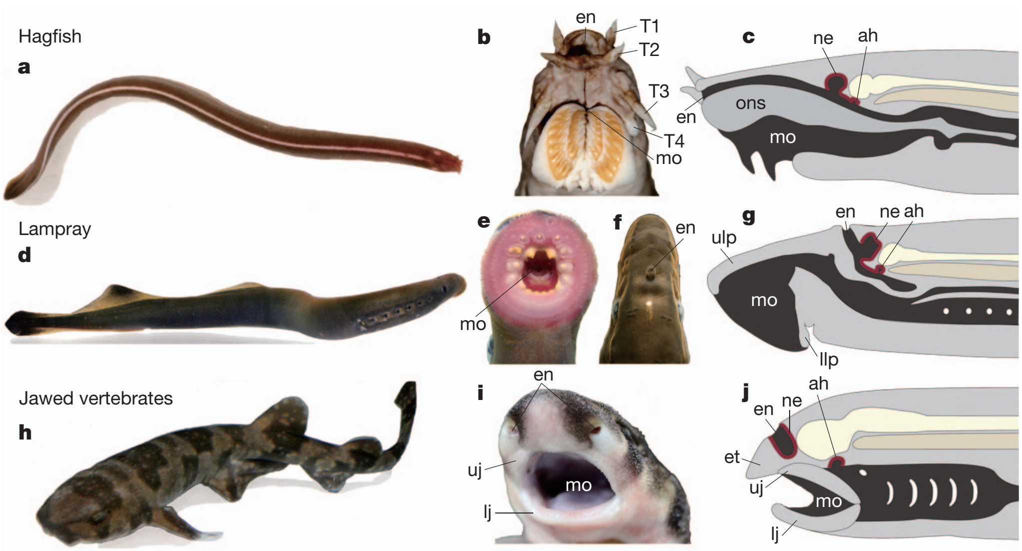


Figure 1 | Cyclostomes and gnathostomes. **a–c**, *Eptatretus burgeri* (hagfish). **a**, Dorsal view. **b**, Ventral view of a hagfish head showing a single nostril (en) and jawless mouth (mo). **c**, A diagram of the head anatomy. **d–g**, *Lethenteron japonicum* (lamprey). **d**, Lateral view. **e**, Ventral view to show the oral funnel surrounding the mouth. **f**, Dorsal view to show the dorsal nostril. **g**, A diagram of the head anatomy. **h**, **i**, *Scyliorhinus torazame* (shark). **h**, Lateral view. **i**, Frontal view to show upper and lower jaws (uj and lj), and paired nostrils (en). **j**, Diagram of the generalized gnathostome head. ah, adenohypophysis; et, ethmoidal region; llp, lower lip; ne, nasal epithelium; ons, oronasohypophyseal septum; T1–4, tentacles 1–4; ulp, upper lip.

[1]Department of Biology, Graduate School of Science, Kobe University, Kobe 657-8501, Japan. [2]Laboratory for Evolutionary Morphology, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan. [3]Marine Research Station, Institute of Cellular and Organismic Biology, Academia Sinica, Yilan 26242, Taiwan. [4]Genome Resource and Analysis Unit, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan.

lampreys and gnathostomes. The endodermal origin of the hagfish adenohypophysis was suggested by limited embryological data based on a few embryos that were fixed in a suboptimal manner[7–10]. Notably, the topographical relationship between the adenohypophysis and the oropharyngeal membrane has remained inconclusive[7,9].

The discrepancy between the morphology and molecular phylogenetics may also be attributable to as-yet-undescribed developmental differences between these agnathans. Therefore, understanding the developmental embryonic patterns in hagfishes and lampreys is critical for comparing the morphology between the two cyclostome groups and characterizing the body plan of hypothetical cyclostome ancestors, thus leading to a better understanding of the earliest evolutionary history of vertebrates. In this study, we incubated time-scaled fertilized eggs of the hagfish *Eptatretus burgeri* and collected serial staged embryos to re-examine the development and germ-layer origin of the adenohypophysis and craniofacial structures (Supplementary Fig. 2 and Supplementary Table 1).

In lampreys and gnathostomes, the adenohypophysis and oronasal region are identified as ectodermal derivatives, arising between the stomodeal ectoderm and pharyngeal endoderm[10]. Similar structures also exist in early hagfish embryos, and the position of the oropharyngeal membrane becomes critical for their germ-layer origins[9,10] (Fig. 2).

Craniofacial development of *E. burgeri*

During the mid-neurula (stage 25) of *E. burgeri*, the presumptive oropharyngeal membrane is anterior to the forebrain primordium (Fig. 2a, b), where the ectoderm and endoderm are juxtaposed with each other (Fig. 2b and Supplementary Fig. 3a). Depression of the stomodeum is enhanced with the growth of the forebrain by the late neurula (stage 28), creating the presumptive oronasohypophyseal cavity anterior to the oropharyngeal membrane (Fig. 2c, d and Supplementary Fig. 3). However, the oronasohypophyseal cavity is secondarily closed from the exterior by a pair of ectodermal folds arising on both sides of the forebrain, which grow rostromedially to fuse with its counterpart (arrows in Fig. 2c, e and Supplementary Fig. 4a, b). This fusion finally forms a hagfish-specific membrane, the secondary oropharyngeal membrane (SOM)[16] (Fig. 2f–i). The previously reported endodermal origin of the hagfish oronasohypophyseal cavity was based on misidentification of the SOM as the oropharyngeal membrane[7,9] (Supplementary Fig. 4c); the hagfish oronasohypophyseal cavity is located rostral to the true oropharyngeal membrane, which arises in a more caudal position (Fig. 2i–l). On the basis of the topographical relationship, the hagfish oronasohypophyseal cavity, including the adenohypophysis (light-blue in three-dimensional images, Fig. 2i–l), is likely to be of ectodermal origin.

To exclude the previously reported possibility that the endodermal pouch beneath the hypothalamus differentiates into the adenohypophysis in hagfishes[7], we observed the development of the hypothalamus-associated structures caudal to the true oropharyngeal membrane. In stage 25 *E. burgeri* embryos, a thickening of the rostralmost part of the pharyngeal endoderm abuts the hypothalamus rostral to the notochord (pcp (prechordal plate) in Fig. 2b), and at stage 26, this endodermal portion expresses *EbHh1* and *EbSix3/6A* (pcp in Fig. 3a, c). On the basis of the topographical and molecular signatures, this thickening probably corresponds to the prechordal plate of other vertebrate embryos[17–19] (Fig. 2i, j and Supplementary Fig. 5). Between stages 35 and 40, the prechordal plate forms a dorsally protruding pouch, reminiscent of Rathke's pouch in gnathostomes (Supplementary

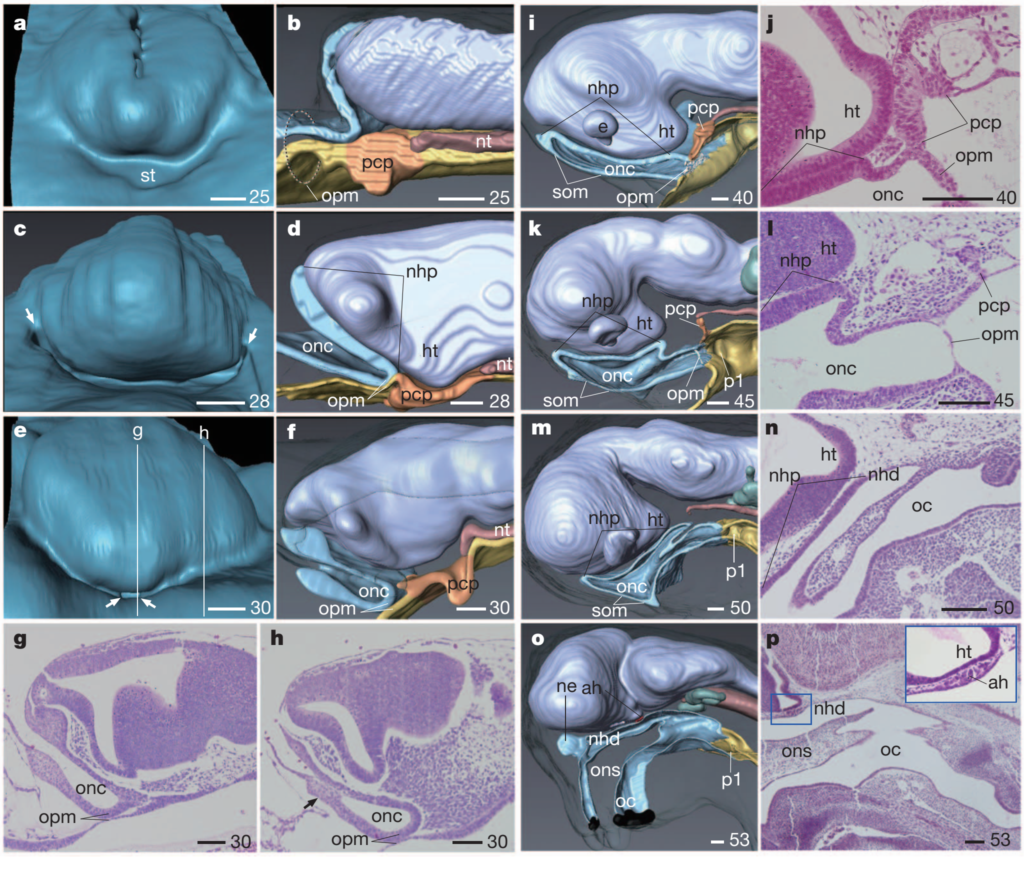


Figure 2 | Craniofacial development of *E. burgeri*. **a–f**, Frontal (**a**, **c**, **e**) and lateral (**b**, **d**, **f**) views. The ectoderm is coloured light blue, the endoderm is coloured yellow, and the mesoendoderm is coloured orange. Arrows indicate the SOM. **g**, **h**, Sagittal sections at the plane represented in **e**. **i–p**, Lateral views (**i**, **k**, **m**, **o**) and sagittal sections (**j**, **l**, **n**, **p**). Inset in **p** shows the differentiated adenohypophysis (ah). e, eye; ht, hypothalamus; nhd, nasohypophyseal duct; nhp, nasohypophyseal plate; nt, notochord; oc, oral cavity; onc, oronasohypophyseal cavity; opm, oropharyngeal membrane; pcp, prechordal plate; p1, pharyngeal pouch 1; st, stomodeum. See Fig. 1 for other abbreviations. Scale bars, 100 μm. Numbers indicate developmental stages.

Fig. 5a–c), adhering to the posterior aspect of the hypothalamus (Fig. 2i, j and Supplementary Fig. 5d–g). This pouch, however, becomes detached from the hypothalamus by stage 45 and disappears by stage 50 without differentiating into any overt structures, with the exception of a pair of lateral chords of cells growing from the pouch (Fig. 2k–n and Supplementary Fig. 5d–i). These chords are reminiscent of the premandibular mesoderm in embryos of other vertebrates[17]; however, they also disappear before stage 45 (Fig. 2i, k). Thus, the rostral endoderm in hagfishes does not participate in the formation of the adenohypophysis.

The ectodermal origin of the hagfish adenohypophysis is consistent with the gene expression patterns. In lampreys and gnathostomes, the molecular signature of the hypothalamic anlage, to which the adenohypophysis is adhered ventrally, is seen as a domain of *Shh* and *Nkx2.1* expression posterior to the *Six3/6*-positive optic chiasma[18,20]. The hypophyseal placode is defined by the expression of *Pitx1/2* and *Lhx3/4* (refs 20, 21). Similarly, in hagfish embryos, *EbHh1* and *EbNkx2.1* (putative orthologues of *Shh* and *Nkx2.1*) are expressed in the presumptive hypothalamus (Fig. 3a, b and Supplementary Fig. 6a, b), caudal to the optic chiasma that expresses *EbSix3/6A*, the putative hagfish orthologue of *Six3/6* (Fig. 3c, d and Supplementary Fig. 6c). Notably, the expression of *EbPitxA* and weak expression of *EbLhx3/4A* (putative *Pitx2* and *Lhx3/4* orthologues in hagfishes) were detected in the ectodermal thickening rostral to the oropharyngeal membrane beneath the hypothalamus (Fig. 3e and Supplementary Fig. 6d–f), which is topographically comparable to the hypophyseal placode of other vertebrates. Thus, the hagfish adenohypophysis is ectodermally derived, apparently based on gene regulations identical to those in gnathostomes and lampreys.

Pan-cyclostome pattern of embryogenesis

Unlike in gnathostomes, the hypophyseal and nasal placodes of the lamprey embryo form a single median complex, the nasohypophyseal plate (NHP)[22] (Supplementary Fig. 7). In hagfish embryos, the putative NHP is an ectodermal thickening beneath the forebrain (nhp in Fig. 2d), which is itself bilaterally encircled by a horseshoe-shaped band of thickened ectoderm, reminiscent of the pan-placodal domain in other vertebrates[15] (Fig. 3g). It was possible to follow the latter band caudally to the otic placode, and it also encompassed the putative precursors of cranial sensory ganglia[23] (Fig. 3g).

The hypophyseal placode was identified in the posterior part of the NHP, whereas its anterior part, or the nasal placode homologue, expressed *EbFgf8/17*, a putative hagfish orthologue of *Fgf8* (Fig. 3f), which is expressed in the nasal placode in gnathostomes and the anterior NHP in lampreys[20,24]. Furthermore, *EbSix3/6A* and *EbSoxB1* were expressed throughout (both anterior and posterior) the NHP of hagfish (Fig. 3c, d and Supplementary Fig. 6c, g, h), consistent with the expression patterns of *Six3/6* and *Sox2/3* in both the nasal and hypophyseal placodes in gnathostomes[15,18]. The NHP takes the form of a single median placode through development (Fig. 3h), and finally differentiates into the adenohypophysis and unpaired nasal epithelium in a single nasohypophyseal duct, as in lampreys (Fig. 2o, p). Thus, the NHP seems to be homologous between hagfishes and lampreys.

The homology of the NHP between hagfish and lampreys implies that the overall craniofacial organization of cyclostomes would also be conserved. At mid-pharyngular stages of cyclostomes, the NHP is rostrocaudally bordered by ventral growths of ectomesenchymal processes, with the anterior nasal process rostral to the NHP and the post-hypophyseal process between the NHP and prechordal plate[22] (Fig. 4a, e and Supplementary Fig. 7). Both in hagfishes and lampreys, the post-hypophyseal process initially arises as a pair of lateral processes at the opening of the oronasohypophyseal cavity (Fig. 4b, c, f, g), which coalesce at the midline to form a shelf behind the NHP, separating the nasohypophyseal duct and oral cavity (Fig. 4d, h). In the late pharyngula of hagfishes, the rostrolateral portion of the post-hypophyseal process divides into four pairs of tentacles (Fig. 4d), whereas the equivalent process of lampreys forms a dorsal oral hood, or the upper lip of the ammocoete larva (Fig. 4h). The peripheral

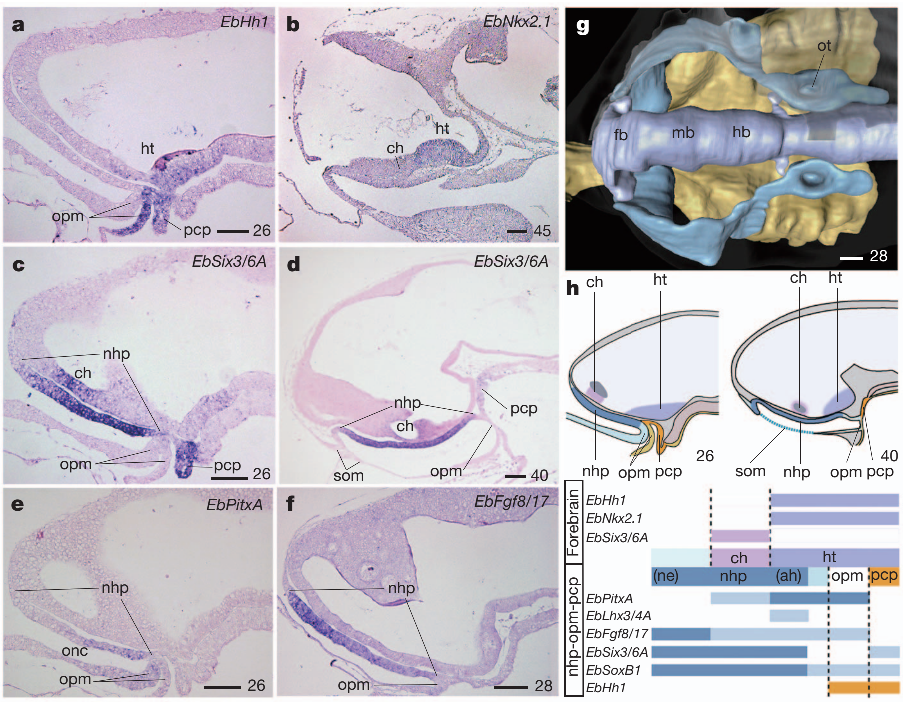


Figure 3 | Embryonic gene expression in *Eptatretus burgeri*. **a**–**f**, Expression of genes involved in the NHP development from stages 26 to 40. **g**, Dorsal view of three-dimensional-reconstructed image. The blue area shows pan-placodal domain. **h**, Schematic diagrams summarizing the gene expression in the embryos. In the graph, the hypothalamic domain is characterized by *EbHh1*, *EbNkx2.1* and *EbSix3/6* expression. Adenohypophysis-like expression (blue bars) is restricted in the ectodermal domain rostral to and including the oropharyngeal membrane (opm). ch, chiasma; fb, forebrain; hb, hindbrain; mb, midbrain; ot, otic placode. See Figs 1 and 2 for other abbreviations. Scale bars, 100 μm.

distribution patterns of the trigeminal nerves are consistent with the above comparison, indicating that a subset of hagfish tentacles (T1, T3, T4) are comparable with the lateral wall of the lamprey upper lip and T2 is comparable with the dorsal medial roof of the upper lip (Supplementary Fig. 8). The oral roof beneath the nasohypophyseal duct is homologous between hagfishes and lampreys[25–27] (Fig. 4l), but it is not comparable to the pattern of jawed vertebrates (Fig. 4i–l and Supplementary Fig. 9).

During the developmental stages when lamprey and hagfish embryos show the most conspicuous resemblance to each other in the oronasal region, other craniofacial portions of the embryos also show similarities, most of which can be recognized as vertebrate-specific features[22]. For example, *EbPax1/9* was expressed in a series of pouches on the lateral side of the pharynx (Supplementary Figs 10a and 11a), whereas *EbTbx1/10A* was expressed in the mesenchymal core in each space between two successive developing pouches (Supplementary Figs 10b and 11a), consistent with the expression patterns of *Pax1/9* and *Tbx1/10* in pharyngeal pouches and the mesodermal core of pharyngeal arches in other vertebrate embryos, including those of lampreys[28,29]. We also noted post-otically developing somites and typical rhombomeric segmental patterns on the hindbrain of hagfishes, as well as the formation of cranial nerve roots on even-numbered rhombomeres (Supplementary Fig. 12a, b). In both of the cyclostomes, the first pharyngeal mesoderm (mandibular mesoderm) and pouches form the muscle anlage of the post-hypophyseal process/velum and velar chamber[30] (Supplementary Fig. 11a–d, f–m). By the late pharyngula, the second pharyngeal pouch of lampreys forms the rostral-most gill pore (Supplementary Fig. 11m), whereas that of hagfishes is derived from pouch 4 because pouches 2 and 3 are degenerated secondarily in hagfishes (Supplementary Fig. 11h–j). Nevertheless, the pharyngeal mesoderm belonging to arches 2–4 remain in the original site, innervated by the equivalent cranial nerves (Supplementary Fig. 12c–g). Thus, the embryonic morphology of the pharyngeal structures in hagfishes also reflects the basic segmental pattern of pharyngeal arches generally found in vertebrates. In the late pharyngula of hagfishes, however, secondary modification of craniofacial structures results in the anatomical patterns unique to hagfishes.

In the late hagfish embryo, the oral cavity and nasohypophyseal duct grow caudally relative to the position of the adenohypophysis (blue and black arrows, Supplementary Fig. 11f–j), and the pharynx elongates posteriorly[31] (white arrow, Supplementary Fig. 11a–j). During these processes, the presumptive nasohypophyseal duct, velum and velar chamber recline inwards to take an oblique orientation unlike that of lamprey, which is positioned more vertically (Fig. 2i, k, m, o and Supplementary Fig. 11f–j). The gill apparatus also shifts caudally to the mid-trunk level (white arrow and sm (somite) in Supplementary Fig. 11a–e). In addition, the posterior end of the nasohypophyseal duct opens into the pharynx by the rupture of the caudal root of the post-hypophyseal process-derived oronasohypophyseal septum (compare Fig. 2m, n to Fig. 2o, p). Thus, unlike the nasohypophyseal duct of lampreys, which ends blindly, the duct in hagfishes opens secondarily into the pharynx. These hagfish-specific morphological modifications are what differentiate adult hagfishes and lampreys at the anatomical level (Supplementary Fig. 1). Alternatively, the overall embryonic features indicate that the larval morphology in lampreys is at a relatively less differentiated state compared with that of hagfishes.

Discussion

Our re-interpretation of the ectodermal adenohypophysis (Fig. 2) refutes the previous hypothesis that the craniofacial structures of hagfishes represent a more ancestral state than those of all other

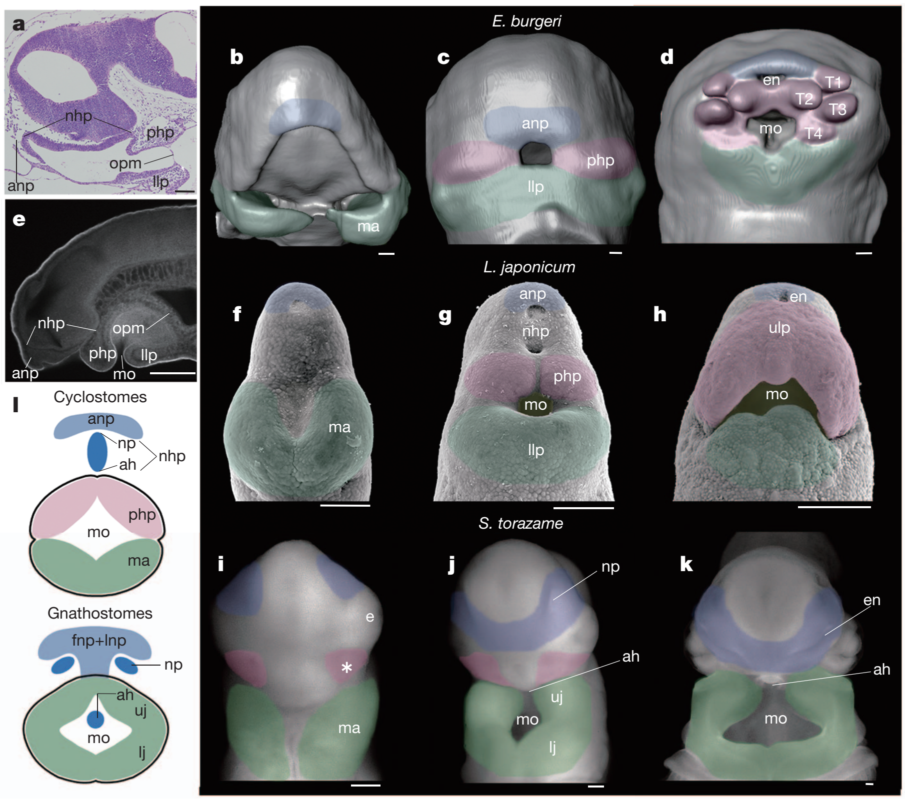


Figure 4 | Comparison of vertebrate heads. a–h, Mid-sagittal sections (**a**, **e**) and ventral views (**b–d**, **f–h**) of *Eptatretus burgeri* (**a**, stage 45; **b**, stage 40; **c**, stage 50; **d**, stage 53; SOM has been removed from **b–d**) and *Lethenteron japonicum* (**e**, stage 26; **f**, stage 23; **g**, stage 26; **h**, stage 27). **i–k**, Ventral views of *Scyliorhinus torazame* (**i**, stage 20; **j**, stage 25; **k**, stage 28). The premandibular domain (asterisk) is comparable to the post-hypophyseal process (php) of cyclostomes. The blue domain (lateral and medial nasal prominences (fnp+lnp)) resembles the anterior nasal process (anp) of cyclostomes (ref. 22). **l**, Generalized patterns of vertebrate heads. ma, mandibular arch; np, nasal placodes. See Figs 1 and 2 for other abbreviations. Scale bars, 100 μm.

vertebrates. This hypothesis, together with the concept of the 'endodermalized head' of hagfish[7,8], was already called into the question by reports that the hagfish adenohypophysis is structurally and functionally similar to that of other vertebrates[32,33]. Also, many of the vertebrate-defining characters once believed to be absent from hagfishes, including lens, spleen, vertebrae and a lateral line system, are now known to develop transiently or in primitive state in hagfish adults and embryos[5,34–36]. In addition, there are even some cyclostome-specific synapomorphies at the morphological level as well, including the rostrally migrating myotomes, velum and tongue apparatus[22,37], that support the monophyly of cyclostomes (Fig. 5). Thus, the morphological status of hagfishes is reasonably consistent with the molecular phylogenetic data[1–5].

Hagfishes and lampreys share embryonic craniofacial features such as the NHP and two processes (and their derivatives) anterior and posterior to the NHP (Fig. 5). This pattern never appears during any developmental stage of living gnathostomes. The adenohypophysis and nasal organ are adjacent to each other anteroposteriorly in the cyclostomes[22] (Fig. 5). In contrast, the adenohypophysis becomes separated from the paired nostril and is secondarily incorporated into the oral ectoderm in gnathostomes (Fig. 5). The possible single nasohypophyseal opening is shared with the basal 'ostracoderms' (stem gnathostomes), the anaspids[6,12,38], such that the NHP can be interpreted as an attribute of the monorhiny of the basal agnathans and thus the ancestral state for vertebrates (Fig. 5).

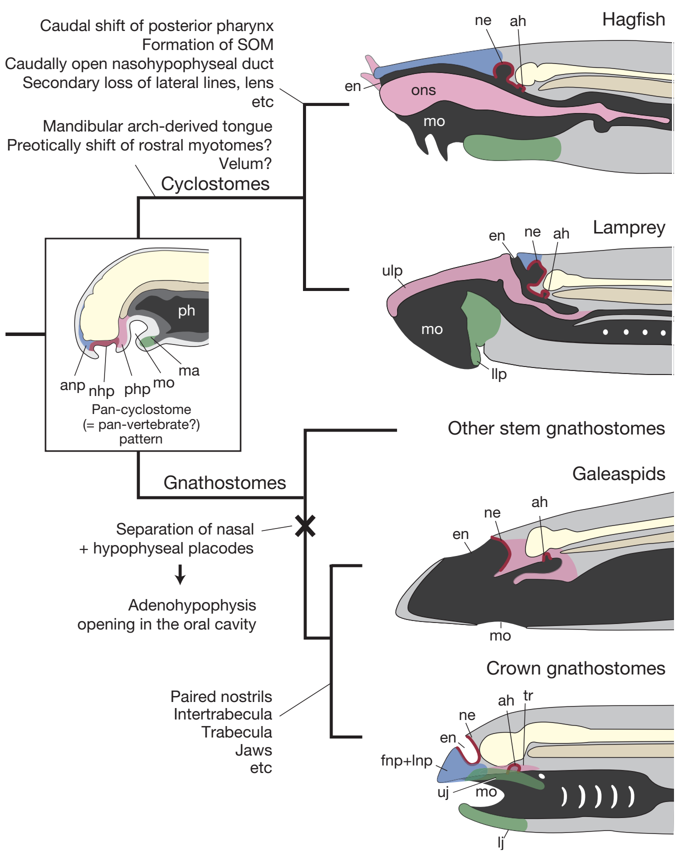


Figure 5 | Evolution of the vertebrate head. A model based on the assumption that the pan-cyclostome pattern (in box) represents a plesiomorphic pattern for the entire vertebrates. The hagfish oronasohypophyseal septum (ons) and the lamprey upper lip (ulp) are both derivatives of the post-hypophyseal process (php). Separation of the NHP into the nasal epithelium (ne) and adenohypophysis (ah) may have led to the pattern of galeaspids[42]. Further bilateral separation of the nasal epithelia is regarded as a prerequisite for acquisition of the trabecula (tr) and upper and lower jaws (uj and lj) in crown gnathostomes. ph, pharynx. See Fig. 1, 2 and 4 for other abbreviations.

After the establishment of the cyclostome developmental pattern, hagfish embryos go through enormous morphological changes, including the caudal shift of pharyngeal structures (Fig. 5 and Supplementary Fig. 11a–j) and opening of the nasohypophyseal duct into the pharynx (Figs 2m–p and 5). For these hagfish-specific, late-embryonic modifications, anatomical comparison of lampreys and hagfishes is difficult, and lampreys and gnathostomes used to be regarded as more closely related to each other[6,11,12]. The autoapomorphic nature of hagfish morphology, especially in the craniofacial region, is consistent with a fossil hagfish, *Myxinikela siroka*, with oropharyngeal structures positioned more rostrally than in the living hagfishes, which can be regarded as representing an intermediate transitional stage[39]. Although the morphology of the nasohypophyseal duct in *M. siroka* remains unclear, the medial subnasal cartilage (T2) beneath the nasohypophyseal duct, at least, is located more dorsally than the lateral labial cartilage (T1)[39]. Thus, hagfishes are likely to be a highly derived lineage of cyclostomes.

Without suitable outgroups, the evaluation of the proposed pan-cyclostome pattern is not easy: if this was the pattern of the common ancestor of cyclostomes, could gnathostomes have evolved by the modification of this same pattern (Fig. 5)? Recent studies have suggested that, even before acquisition of the jaw, major components of the vertebrate head were re-organized in gnathostome lineages[40]. In the lineage of gnathostome, evolutionary polarity from single to double nostrils is apparent. If the monorhiny in stem gnathostomes shared the same origin as that of cyclostomes, that would be highly indicative of NHP in their embryos. In this context, two informative groups of fossil agnathans are galeaspids and osteostracans, the endoskeletons of which are well ossified. In particular, the cranial morphology of osteostracans has traditionally been compared with the lamprey with its single nasohypophyseal duct that opens dorsally[41]. However, a recent observation of galeaspids, a lineage basal to the osteostracans[38], has suggested that their adenohypophysis was separated from the nasal sac and opened into the oral cavity, as in jawed vertebrates[42]. If this animal truly represents a transition from mono- to diplorhiny states, it would represent an intermediate condition towards the crown gnathostomes[6,42] (Fig. 5). Thus, the lamprey-like morphology of osteostracans would represent a secondary condition. Even so, the embryonic development of galeaspids may have initially passed through the cyclostome pattern, yet with a widely expanded NHP that preceded its break-up into nasal and hypophysial placodes, as seen in jawed vertebrates. Thus, the pan-cyclostome pattern may possibly represent a symplesiomorphic ancestral state of the craniofacial development of the vertebrates. If so, the major part of the gnathostome craniofacial evolution could possibly be seen as the loss and modification of the cyclostome-like ancestral patterns, especially characterized by the acquisition of diplorhiny, upon which the evolution of the prechordal cranium and jaw depended[40,42] (Fig. 5).

METHODS SUMMARY

Eptatretus burgeri embryos were collected as described previously[43], staged according to ref. 44 (Supplementary Table 1), and fixed with Serra's fixatives or 4% paraformaldehyde. Sections were made by Kawamoto's protocol (Section Lab Co. Ltd). Haematoxylin and eosin staining followed standard protocols. Images were recorded with a DP70 digital camera (Olympus Corporation) attached to a light microscope and reconstructed with Avizo (Visualization Sciences Group).

Full Methods and any associated references are available in the online version of the paper.

Received 2 October; accepted 19 November 2012.
Published online 19 December 2012.

1. Mallatt, J. & Sullivan, J. 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. *Mol. Biol. Evol.* **15,** 1706–1718 (1998).
2. Kuraku, S., Hoshiyama, D., Katoh, K., Suga, H. & Miyata, T. Monophyly of lampreys and hagfishes supported by nuclear DNA-coded genes. *J. Mol. Evol.* **49,** 729–735 (1999).
3. Takezaki, N., Figueroa, F., Zaleska-Rutczynska, Z. & Klein, J. Molecular phylogeny of early vertebrates: monophyly of the agnathans as revealed by sequences of 35 genes. *Mol. Biol. Evol.* **20,** 287–292 (2003).

4. Kuraku, S. Insights into cyclostome phylogenomics: pre-2R or post-2R. *Zoolog. Sci.* **25,** 960–968 (2008).
5. Heimberg, A. M., Cowper-Sal-lari, R., Semon, M., Donoghue, P. C. & Peterson, K. J. microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. *Proc. Natl Acad. Sci. USA* **107,** 19379–19383 (2010).
6. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, 1996).
7. Gorbman, A. Early development of the hagfish pituitary gland: evidence for the endodermal origin of the adenohypophysis. *Am. Zool.* **23,** 639–654 (1983).
8. Gorbman, A. & Tamarin, A. In *Evolutionary Biology of Primitive Fishes* (eds Foreman, R. E., Gorbman, A., Dodd, J. M. & Olsson, R.) 165–185 (Plenum, 1985).
9. Wicht, H. & Tusch, U. in *The Biology of Hagfish* (eds Jørgensen, J. M., Lomholt, J. R., Weber, R. E. & Molte, H.) 431–451 (Chapman & Hall, 1998).
10. Soukup, V., Horácek, I. & Cerny, R. Development and evolution of the vertebrate primary mouth. *J. Anat.* http://dx.doi.org/10.1111/j.1469-7580.2012.01540.x (16 July 2012).
11. Forey, P. & Janvier, P. Agnathans and the origin of jawed vertebrates. *Nature* **361,** 129–134 (1993).
12. Gess, R. W., Coates, M. I. & Rubidge, B. S. A lamprey from the Devonian period of South Africa. *Nature* **443,** 981–984 (2006).
13. Marinelli, W. & Strenger, A. *Vergleichende Anatomie und Morphologie der Wirbeltiere 2. Myxine Glutinosa* (Franz. Deuticke, 1956).
14. Candiani, S., Holland, N. D., Oliveri, D., Parodi, M. & Pestarino, M. Expression of the amphioxus *Pit-1* gene (*A mphiPOU1F1/Pit-1*) exclusively in the developing preoral organ, a putative homolog of the vertebrate adenohypophysis. *Brain Res. Bull.* **75,** 324–330 (2008).
15. Schlosser, G. Evolutionary origins of vertebrate placodes: insights from developmental studies and from comparisons with other deuterostomes. *J. Exp. Zool. B* **304,** 347–399 (2005).
16. von Kupffer, C. *Studien zur Vergleichenden Entwicklungsgeschichte des Kopfes der Kranioten. Heft 4: Zur Kopfentwicklung von Bdellostoma* 1–86 (Lehmann, 1900).
17. Adachi, N. & Kuratani, S. Development of head and trunk mesoderm in the dogfish, *Scyliorhinus torazame*. I. Embryology and morphology of the head cavities and related structures. *Evol. Dev.* **14,** 234–256 (2012).
18. Jean, D., Bernier, G. & Gruss, P. *Six6* (*Optx2*) is a novel murine *Six3*-related homeobox gene that demarcates the presumptive pituitary/hypothalamic axis and the ventral optic stalk. *Mech. Dev.* **84,** 31–40 (1999).
19. Sugahara, F. *et al.* Involvement of Hedgehog and FGF signalling in the lamprey telencephalon: evolution of regionalization and dorsoventral patterning of the vertebrate forebrain. *Development* **138,** 1217–1226 (2011).
20. Uchida, K., Murakami, Y., Kuraku, S., Hirano, S. & Kuratani, S. Development of the adenohypophysis in the lamprey: evolution of epigenetic patterning programs in organogenesis. *J. Exp. Zool. B* **300,** 32–47 (2003).
21. Charles, M. A. *et al. PITX* genes are required for cell survival and *Lhx3* activation. *Mol. Endocrinol.* **19,** 1893–1903 (2005).
22. Kuratani, S. Evolution of the vertebrate jaw from developmental perspectives. *Evol. Dev.* **14,** 76–92 (2012).
23. Kuratani, S. & Ota, K. G. Primitive versus derived traits in the developmental program of the vertebrate head: views from cyclostome developmental studies. *J. Exp. Zool. B* **310,** 294–314 (2008).
24. Maruoka, Y. *et al.* Comparison of the expression of three highly related genes, *Fgf8, Fgf17* and *Fgf18*, in the mouse embryo. *Mech. Dev.* **74,** 175–177 (1998).
25. Koyama, H., Kishida, R., Goris, R. C. & Kusunoki, T. Organization of sensory and motor nuclei of the trigeminal nerve in lampreys. *J. Comp. Neurol.* **264,** 437–448 (1987).
26. Nishizawa, H., Kishida, R., Kadota, T. & Goris, R. C. Somatotopic organization of the primary sensory trigeminal neurons in the hagfish, *Eptatretus burgeri. J. Comp. Neurol.* **267,** 281–295 (1988).
27. Song, J. & Boord, R. L. Motor components of the trigeminal nerve and organization of the mandibular arch muscles in vertebrates: phylogenetically conservative patterns and their ontogenetic basis. *Acta Anat.* **148,** 139–149 (1993).
28. Ogasawara, M., Shigetani, Y., Hirano, S., Satoh, N. & Kuratani, S. *Pax1/Pax9*-related genes in an agnathan vertebrate, *Lampetra japonica*: expression pattern of *LjPax9* implies sequential evolutionary events toward the gnathostome body plan. *Dev. Biol.* **223,** 399–410 (2000).
29. Tiecke, E. *et al.* Identification and developmental expression of two *Tbx1/10*-related genes in the agnathan *Lethenteron japonicum. Dev. Genes Evol.* **217,** 691–697 (2007).
30. Strahan, R. The velum and the respiratory current of *Myxine. Acta Zool.* **39,** 227–240 (1958).
31. Stockard, C. R. The development of the mouth and gills in *Bdellostoma stouti. Am. J. Anat.* **5,** 481–517 (1906).
32. Sower, S. A., Freamat, M. & Kavanaugh, S. I. The origins of the vertebrate hypothalamic–pituitary–gonadal (HPG) and hypothalamic–pituitary–thyroid (HPT) endocrine systems: new insights from lampreys. *Gen. Comp. Endocrinol.* **161,** 20–29 (2009).
33. Uchida, K. *et al.* Evolutionary origin of a functional gonadotropin in the pituitary of the most primitive vertebrate, hagfish. *Proc. Natl Acad. Sci. USA* **107,** 15832–15837 (2010).
34. Stockard, C. R. The embryonic history of the lens in *Bdellostoma stouti* in relation to recent experiments. *Am. J. Anat.* **6,** 511–515 (1906).
35. Ota, K. G., Fujimoto, S., Oisi, Y. & Kuratani, S. Identification of vertebra-like elements and their possible differentiation from sclerotomes in the hagfish. *Nature Commun.* **2,** 373 (2011).
36. Wicht, H. & Northcutt, R. G. Ontogeny of the head of the Pacific hagfish (*Eptatretus stouti*, Myxinoidea): development of the lateral line system. *Phil. Trans. R. Soc. Lond. B* **349,** 119–134 (1995).
37. Yalden, D. W. Feeding mechanisms as evidence for cyclostome monophyly. *Zool. J. Linn. Soc.* **84,** 291–300 (1985).
38. Janvier, P. Early jawless vertebrates and cyclostome origins. *Zoolog. Sci.* **25,** 1045–1056 (2008).
39. Bardack, D. First fossil hagfish (Myxinoidea): a record from the Pennsylvanian of Illinois. *Science* **254,** 701–703 (1991).
40. Davis, S. P., Finarelli, J. A. & Coates, M. I. *Acanthodes* and shark-like conditions in the last common ancestor of modern gnathostomes. *Nature* **486,** 247–250 (2012).
41. Stensiö, E. A. *The Downtonian and Devonian Vertebrates of Spitsbergen. Part 1: Family Cephalaspidae* (Arno, 1927).
42. Gai, Z., Donoghue, P. C., Zhu, M., Janvier, P. & Stampanoni, M. Fossil jawless fish from China foreshadows early jawed vertebrate anatomy. *Nature* **476,** 324–327 (2011).
43. Ota, K. G., Kuraku, S. & Kuratani, S. Hagfish embryology with reference to the evolution of the neural crest. *Nature* **446,** 672–675 (2007).
44. Dean, B. On the embryology of *Bdellostoma stouti*. A genera account of myxinoid development from the egg and segmentation to hatching. *Festschrift zum 70ten Geburstsstag Carl von Kupffer* 220–276 (Gustav Fischer, 1899).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank O. Kakitani of Shimane Fishermen's Union and K. Kayano of Sekikatsu Inc. for hagfish sample collection; T. Kawamoto for technical advice on paraffin sectioning; I. Kamimura for advice on the Avizo technique; T. Imai and Y. Ashida for advice on the technique for section *in situ* hybridization; and M. Takechi, T. Hirasawa, N. Adachi, Y. Murakami, S. Yonemura, K. Misaki, K. Yasui, S.-i. Aota, M. Nozaki and H. Koyama for their technical support and advice. We also thank M. Tanaka and K. Onimaru for providing the *LjTbx1/10A* plasmid, and H. Nagashima, R. Ladher, F. Sugahara and J. Pascual Anaya for critical reading of the manuscript.

Author Contributions Y.O. performed sample collection, maintenance of aquarium tanks, histological preparation and three-dimensional reconstructions. Y.O. and S.F. performed the molecular cloning of *EbPitxA*, *EbSix3/6A*, *EbFgf8/17A*, *EbHh1*, *EbTbx1/10A*, *EbSoxB1*, *EbLhx3/4A* and *EbNkx2.1* genes and *in situ* hybridization. S. Kuraku performed the molecular evolutionary analysis. Y.O. and S. Kuratani wrote the first draft of the manuscript. K.G.O., S. Kuraku and S. Kuratani wrote the final version of the manuscript. All of the authors discussed the results and commented on the manuscript.

Author Information Sequences for *EbPitxA*, *EbSix3/6A*, *EbFgf8/17*, *EbHh1*, *EbTbx1/10A*, *EbSoxB1*, *EbLhx3/4A* and *EbNkx2.1* from *E. burgeri* are deposited in DDBJ/GenBank/EMBL under accession numbers AB703678–AB703682, AB729075–AB729076, and AB747372. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S. Kuratani (saizo@cdb.riken.jp).

METHODS

Sample collection and aquarium maintenance of hagfish. *Eptatretus burgeri* embryos were collected as described previously[43], staged according to ref. 44 (Supplementary Table 1), and fixed with Serra's fixatives or 4% paraformaldehyde.

cDNA cloning and sequencing. *EbPitxA*, *EbSix3/6A*, *EbFgf8/17*, *EbHh1*, *EbTbx1/10A*, *EbSoxB1*, *EbLhx3/4A* and *EbNkx2.1A* fragments were amplified by degenerate PCR with reverse transcription and isolated by the TOPO TA cloning kit dual promoter (Invitrogen). The sequence data were submitted to the DDBJ database (AB703678–AB703682, AB729075–AB729076, and AB747372). To confirm the orthology of newly identified hagfish genes, homologous sequences were retrieved from the NCBI Protein database and Ensembl Genome Browser, and multiple sequence alignments were generated using the multiple alignment program MAFFT[45], followed by visual inspection and manual curation. On the basis of these alignments, molecular phylogenetic trees were inferred with the neighbour-joining and maximum-likelihood methods (data not shown).

Histology of hagfish embryos. To avoid distortion of histological sections and particularly to preserve the oropharyngeal membrane, Kawamoto's Film Method was employed using the Paraffin Section Preparation kit (Section Lab Co. Ltd; see http://section-lab.jp/English.htm). The adhesive side of the Paraffin Transfer Film (fine type) was attached firmly onto the cut surface of the paraffin block, and each section was cut slowly at 6–8 μm thick with a microtome. One side of the cut 'film-sections' was firmly placed onto a glass slide with the sectioned side down, and the slide was bathed in water at 45 °C for 10 s. The films were then pressed and attached on the glass slide with a rubber roller, and dried on the hot plate at 50 °C for 24 h to transfer the embryonic tissues onto the slides. The adhesive compound and paraffin were removed with xylene for 1–3 days before histological, immunohistochemical and *in situ* hybridization procedures. Images were recorded with a DP70 digital camera (Olympus Corporation) attached to a light microscope and reconstructed with Avizo (Visualization Sciences Group).

***In situ* hybridization of hagfish.** *In situ* hybridization was performed either by using a manual standard protocol or a Ventana automated instrument (Roche Diagnostics). In the standard protocol, serial sections were fixed for 10 min in 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS) at room temperature, washed twice in PBS, treated with ProK in 0.01 M Tris buffer for 10 min, and fixed again for 10 min in 4% PFA at room temperature. After rinsing twice in PBS, the sections were incubated with 0.25% acetic anhydride and 0.1 M triethanolamine (pH 8), washed in PBS, air dried, and hybridized at 51 °C for 16–20 h. The sections were then washed in 5× SSC at 55 °C, treated with 50% formamide in 2× SSC at 60 °C for 20 min, washed in 2× SSC and twice in 0.2× SSC at 60 °C each for 20 min. After blocking with 1.5% blocking reagent (Roche Diagnostics) in 0.1 M Tris buffer with 0.15 M NaCl (pH 7.6), the sections were incubated with alkaline phosphatase-conjugated anti-DIG antibody (Roche Diagnostics). After the final washes with Tris buffer, positive cells were stained purple with nitroblue tetrazolium salt (NBT) and 5-bromo-4-chloro-3-indolyl phosphate toludinium salt (BCIP). When using the Ventana instrument, detection of signals and counterstaining were performed by using the BlueMap NBT/BCIP substrate kit and a nuclear fast red-equivalent reagent, ISH RED, was applied (Roche Diagnostics), as described previously[43].

Sample collection and aquarium maintenance of lampreys. Mature male and female lampreys, *Lethenteron japonicum*, were collected from the Miomote River in Niigata and the Shiribetsu River in Hokkaido, Japan, during the breeding season (early June). Eggs were fertilized artificially and incubated in 10% Steinberg's solution at 16–23 °C (ref. 46). Lamprey embryos were staged according to Tahara's staging of *Lethenteron reissneri*[47], a brook lamprey species closely related to *L. japonicum*[19], and fixed in 4% PFA/PBS overnight at 4 °C. The embryos were gradually dehydrated in a graded series of methanol/PBS. Finally, they were stored in 100% methanol at −20 °C.

Histology of lamprey embryos. Fixed lamprey embryos stored in methanol (−20 °C) were placed in dimethylsulphoxide (DMSO) and methanol (1:1). After washing with Tris-HCl-buffered saline (TST) containing 5% DMSO (TSTd), the embryos were blocked with aqueous 1% periodic acid and 5% non-fat dry milk in TSTd (TSTM). The embryos were incubated in the primary antibody (anti-acetylated tubulin (Sigma) diluted 1:1,000 in TSTM) for 2–4 days at room temperature. After washing with TST, the samples were incubated with Alexa Fluor 488 goat anti-mouse IgG (H+L) secondary antibody (diluted 1:1000 in TSTM; Molecular Probes). After the final wash in TSTd, the embryos were dehydrated and clarified with a 1:2 mixture of benzyl alcohol and benzyl benzoate, in which cell and nucleic membrane were counterstained with membrane-bound fluorescent dye (Cell Tracker CM-DiI; Invitrogen) to be observed with a confocal microscope (LSM710; Zeiss, Inc.).

Scanning electron microscopy of lamprey embryos. Embryos were fixed with 2.5% glutaraldehyde in 0.01 M PBS, washed in 0.01 M phosphate buffer, and post-fixed with 4% osmium tetraoxide. The embryos were then rinsed in phosphate buffer, dehydrated with a graded series of ethanol, and freeze-dried with *t*-butyl alcohol. The embryos were placed on an aluminium stub, sputter-coated with gold–palladium alloy, and viewed with a scanning electron microscope (JSM-5800, JEOL), as described previously[48].

***In situ* hybridization of lamprey.** Whole-mount *in situ* hybridization of lamprey, *L. japonicum*, was performed as described previously[49], with minor modifications. Hybridization and post-hybridization washes were performed at 65 °C to avoid nonspecific cross-hybridization among paralogues.

Sample collection and aquarium maintenance of shark. Fertilized eggs of *Scyliorhinus torazame* were obtained from adult animals that were bred at 16 °C in seawater tanks. Shark embryos were staged according to Ballard's staging of *Scyliorhinus canicula*, a species closely related to *S. torazame*[17,50].

45. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33,** 511–518 (2005).
46. Steinberg, M. A nonnutrient culture medium for amphibian embryonic tissues. *Year B. Carnegie Inst. Wash.* **56,** 347–348 (1957).
47. Tahara, Y. Normal stages of development in the lamprey, *Lampetra reissneri* (Dybowski). *Zoolog. Sci.* **5,** 109–118 (1988).
48. Horigome, N. *et al.* Development of cephalic neural crest cells in embryos of *Lampetra japonica*, with special reference to the evolution of the jaw. *Dev. Biol.* **207,** 287–308 (1999).
49. Takio, Y. *et al. Hox* gene expression patterns in *Lethenteron japonicum* embryos: insights into the evolution of the vertebrate Hox code. *Dev. Biol.* **308,** 606–620 (2007).
50. Ballard, W. W., Mellinger, J. & Lechenaut, H. A series of normal stages for development of *Scyliorhinus canicula*, the lesser spotted dogfish (Chondrichthyes: Scyliorhinidae). *J. Exp. Zool.* **267,** 318–336 (1993).

CAREERS

TURNING POINT Neuroscientist's successes lead to post in his native country **p.261**

NATUREJOBS BLOG News and commentary on scientific careers **go.nature.com/ielkkf**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

TIM ELLIS/IKON IMAGES/CORBIS

GRADUATE STUDENTS

Structured study

European institutions are introducing organized doctoral programmes that broaden students' education.

BY AMBER DANCE

Kathy Kennedy faced a decision six months into her PhD. Having spent time in various labs studying parasitic liver flukes and the intestinal bacterium *Clostridium difficile*, she had to choose where to spend the next three-and-a-half years while completing her thesis. The rotations helped her not only to make an educated decision about which research group to join, but also to meet colleagues and learn extra skills.

Five or ten years ago, her experience would have been unusual in Europe, where doctoral students would find a mentor, then immediately join his or her research group. They rarely had classes or a choice of research topics. "I went into the lab until I was done," recalls Christine Loscher, who directs the Bioanalysis and Therapeutics Structured PhD Programme (BioAT) at Dublin City University, and is Kennedy's chosen adviser.

A doctorate in Europe traditionally involved a one-on-one relationship between master and apprentice. In Germany, for example, the word for that influential mentor is *Doktorvater* — 'doctor father'. Pursuing related interests and coursework was not generally on the agenda. PhD holders usually went on to make their careers in academia, so the *Doktorvater* was training his or her replacement.

Kennedy's experience reflects a new kind of doctorate. A structured PhD generally includes the research training of a conventional degree, together with other elements that keep students on track and help to prepare them for a variety of post-PhD careers. Students have multiple advisers and administrators to guide them, receive formal training in scientific and soft skills and have opportunities for travel and interdisciplinary study. In principle, they receive a broader education than any single supervisor could provide.

NEW-LOOK DOCTORATES

The rise of structured PhDs has been rapid. In 2007, just over one-quarter of European institutions had established structured doctoral programmes; by 2010, it was almost two-thirds, according to the European University Association (EUA) in Brussels[1]. And whereas in 2007 almost half the institutions offered formalized classwork — usually taking the form of short modules that ran for a day or two — 72% did so in 2010.

In many ways, Europe's move towards structured PhD programmes is making the ▶

continent's doctoral education more like that in the United States, where students apply to programmes rather than to principal investigators, and undertake coursework and rotations before beginning their theses. According to a survey[2] of 160 institutions released in October 2012, more than 90% of North American universities offer structured PhDs. But differences remain: US PhDs tend to take longer — most students graduate in seven years, compared with four in the European Union (EU) — because they include the training that European students typically get in a separate master's degree. EU programmes also tend to put more emphasis on broad skill sets likely to be useful in non-academic positions.

"In the traditional model, it was just assumed that if you're a good researcher, you're a good supervisor."
Thomas Jørgensen

In Europe, the implementation of 'structure' varies widely. In some places, such as France and the United Kingdom, structured programmes are already the norm. Others cling to the conventional model or have not yet implemented structured PhDs. Even within a nation, some institutions or departments may offer structured programmes whereas others do not.

GATHERING SKILLS

Now in the second year of her PhD, Kennedy has taken classes in presentation skills, statistics and scientific writing, as well as modules more specific to her research topic: the immune system's recognition of *Clostridium* surface proteins. BioAT is a joint effort between six Irish institutions, and Kennedy will interact with colleagues in other disciplines, institutions and countries. She will meet formally with Loscher and a second adviser twice a year to make sure that her work is on track. She will also have the chance to study intellectual property and commercialization — which will come in handy if she, like many PhD students, ends up pursuing a career in industry. "You become a more well-rounded student because all these different areas are covered," she says.

Candidates who have trained in structured programmes, and taken classes on topics such as patent searches and intellectual property, can hit the ground running in an industry job, says Declan Moran, a manager at Ipsen, a pharmaceutical company in Dublin.

And although such extra course requirements may mean a greater commitment to the classroom, one's research project need not be neglected, say advocates. Ideally, the classes and programme together provide a framework for students' main studies. "This does not diminish the PhD itself as an original piece of research," says Conor O'Carroll, research director at the Irish Universities Association in Dublin. "It's to enhance the traditional PhD."

There is some early evidence that it is working. O'Carroll conducted surveys of a total of 1,455 PhD trainees across seven Irish universities in 2008, 2009 and 2010 (ref. 3). Even in that short period, students' experiences changed as more of them entered structured programmes. For example, the number of students who had received detailed timetables at the start of their PhDs rose by 20% between 2008 and 2010. Students also became 25% more likely to discuss their research projects in detail with their advisers at the start of their programmes, and there was a 20% increase in student satisfaction with the availability of their mentors. The number of students with more than one supervisor rose by 18%. All these factors contribute to a happier student, says O'Carroll.

Structured programmes also boosted candidates' scientific output, he found: their students were 6% more likely to publish a paper than were those in conventional training. They were also 5% more likely to present their work at an international meeting.

UNEXPECTED BENEFIT

Students come out of structured programmes with a broad set of skills — some of which they may not have known they needed. When she began her PhD in biochemistry at the Free University of Berlin, Agnieszka Denkis was fairly sure that she wanted an academic career. But three years later, she is considering entering industry to do research with real-world applications. She doesn't have to look far for information on how to make the transition — her programme offers workshops on industrial careers.

"You have to like what you're doing. You're going to be married to it."
Sarah-Jo Sinnott

Denkis is also developing transferable skills outside official modules. She has helped to organize a symposium, getting practice in fund-raising and inviting presenters. And through teaching, she has worked on her public speaking. Trainees in conventional PhDs can also seek out these opportunities, but structured programmes organize them and make them easier for students to find.

Students in structured PhD programmes are generally not dependent on a single supervisor: Sarah-Jo Sinnott, who is in her third year of the Health Services Research programme at University College Cork in Ireland, has four. "I seem to collect them," she jokes. When she started her studies, she was assigned one mentor, who studies dental public health. As she developed her thesis topic, a study of new insurance co-payments for prescriptions in Ireland, she recruited a lecturer in pharmacy studies and two economists. Her programme is a multi-institutional effort, so she can receive guidance from faculty members at University College Cork, Trinity College Dublin and the Royal College of Surgeons in Ireland, also in Dublin.

Having a diverse committee means that trainees have someone to go to if they fall out with their primary advisers. They can also get multiple references after they graduate.

In some programmes, the supervisors, too, receive structured training. "In the traditional model, it was just assumed that if you're a good researcher, you're a good supervisor," says Thomas Jørgensen, senior programme manager for the Council for Doctoral Education at the EUA. These days, by contrast, Irish faculty members can attend workshops in mentoring. Swedish advisers are often required to pass a test on rules such as how much time they must set aside for advising. Other new supervisors may simply share a coffee with a more seasoned mentor to get some advice, says Jørgensen.

Through classes, rotations and social events that bring together students working on different research topics, structured programmes facilitate networking that might not otherwise be available to students. That enhanced interaction pays dividends, says Eric Haertel, a second-year PhD student in the Molecular Life Sciences programme at the Swiss Institute of Technology (ETH) in Zurich. Trainees from ETH and the University of Zurich, which is a partner in the programme, get together for barbecues and to drink tea, and attend an annual retreat with student presentations and a party. Without the programme, Haertel wouldn't know anybody at the university, he says. The most valuable element of those social events is chatting with other students about his and their work, and solving problems together. And simply knowing who else works on the same topic can be useful, says Haertel, when he runs out of reagents and needs to borrow something.

The networking of PhD candidates helps their advisers, too, because students can bring back valuable information from other labs, or can start collaborations, says Loscher. For example, during one of her rotations, Kennedy performed luciferase assays, which measure the expression of genes linked to the firefly enzymes that produce light; now she can teach the technique to the rest of Loscher's lab group.

ACROSS BORDERS AND SECTORS

Many structured programmes promote networking across national and disciplinary borders. Ilaria Alborelli entered a programme at the University of Basel in Switzerland with an eye towards a project in fruitfly development.

But she broadened her experience through rotations in neurobiology and stem-cell labs. The programme also includes funding for international travel. "It's good to experience research in different environments," says Alborelli. Making it easy for students to travel also helps them to reach the best labs for their research, adds O'Carroll.

The process of applying for a structured PhD varies with the programme. In the conventional model, trainees found a professor to take them on and joined that lab. In some universities, this still happens, but the principal investigator then encourages the student to join an appropriate programme.

In other cases, applicants approach the programme first, and faculty members jointly decide whether their qualifications merit an interview. During the interview visit, applicants meet several faculty members and look for a good match; if a professor and student want to work together, then the student is admitted. And in some programmes, students do not even decide on a mentor until they have started their degrees and completed their rotations. This 'group application' process helps faculty members too, says O'Carroll. Instead of picking through individual students, they can let the programme select the best candidates.

Structured studies may not be the best fit for everyone, says Sarah Weigelt, a postdoc at the University of Münster in Germany, who will start a job as a junior professor at the University of Bochum in April. "It's quite dependent on the personality," she says. Some trainees thrive in a clearly defined curriculum, but others know exactly what research they want to do, and Weigelt wouldn't want them to be delayed by structured coursework. She would prefer to see universities offer a range of PhD options so that students can choose their ideal path.

Although picking the right sort of programme matters, the key factor in success may still be finding the right research project. "You have to like what you're doing," says Sinnott. "You're going to be married to it." ■

Amber Dance *is a freelance science writer in Los Angeles, California.*

1. Sursock, A. *et al. Trends 2010: A decade of change in European Higher Education* (EUA, 2010).
2. *TRANS-DOC Survey on PhD Programme Structures and Administration in Europe and North America* Presented 15 October 2012 in Brussels; available at go.nature.com/sm6lp9.
3. O'Carroll, C. *et al.* The PhD in Europe: Developing a System of Doctoral Training That Will Increase the Internationalisation of Universities. In *European Higher Education at the Crossroads* (eds Curaj, A., Scott, P., Vlasceanu, L. & Wilson, L.) 461–484 (Springer, 2012).

TURNING POINT
Rui Costa

Rui Costa, a neuroscientist at the Champalimaud Center for the Unknown in Lisbon, accepted his latest accolade last October: a US$15,000 Young Investigator Award from the Society of Neuroscience in Washington DC, for his innovative approach in detailing the brain circuits involved in learning. Trained as a veterinary surgeon, Costa traces his success to a 1997 New Year's resolution to follow his research muse.

Has animal behaviour always interested you?
Yes; as a kid, I loved watching natural-history television programmes. In my last year of veterinary school, I did a research internship at the Swedish University of Agricultural Sciences in Skara, in which I studied suckling behaviour in cows and dolphins. I wanted to understand the origins of behaviour and realized that I would have to study the brain.

What do you consider the most pivotal moment in your career?
I worked as a dairy-cow vet for one year after getting my doctorate of veterinary medicine at the Technical University of Lisbon. But I knew that I had to follow my passion for research. On New Year's Eve of 1997, I quit. I had no job and no salary, but I wanted so badly to be a researcher that there was no question that I would find a way to make it work. It sounds crazy, but putting yourself in a position of not going back frees you to pursue your passion.

How did you end up in the United States at the University of California, Los Angeles (UCLA)?
I found a government-sponsored programme at the University of Porto in Portugal, in which students took a year of classes and could then do their research at any university that would fund them. I decided to contact memory researcher Alcino Silva, who was moving to UCLA.

Why did you choose Silva?
He had done the first genetic manipulation of the mouse brain. I believed that looking at brain circuits was the best way to study genetics and behaviour, so I e-mailed him. It wasn't my technical skills that convinced him to take me on, so it must have been my passion for the research.

Were the experiments that you wanted to conduct technologically feasible?
Not really. I have always wanted to image brain activity while genetically manipulated animals were learning and behaving, but the technology didn't exist then. Even today, we can still image only the surface of the brain. But I knew that Miguel Nicolelis, a neurobiologist at Duke University in Durham, North Carolina, was recording many primate neurons at once using hundreds of electrodes. Looking for a postdoc position, I wrote to him and said that I would love to modify this technique to record activity in mouse brains. He immediately replied that he had had the same idea, and called a few minutes later. He used a phrase in Brazilian Portuguese that means, "Let's go ahead, because people are coming from behind".

Did you always intend to return to Portugal?
I married an American and I liked being in the United States. Yet I missed my family in Portugal and I felt that because Portugal had given me this opportunity, I would one day return. However, I didn't want to just go back and not do science. In 2004, António de Sommer Champalimaud, a Portuguese entrepreneur, died and willed roughly one-third of his fortune to launching the Champalimaud Foundation. The foundation decided to start a neuroscience programme at the Champalimaud Center for the Unknown, with two other researchers and me. In 2007, we shook hands to seal the deal.

How has the experience been so far?
Everything has gone smoothly. The building has an excellent view of the sea and a huge open lab space. We were able to recruit great talent from around the world and have grown to 15 labs, with 130 people. When I arrived, I got a big grant from the European Research Council; last year, I received a 5-year International Early Career Scientist award from the Howard Hughes Medical Institute in Chevy Chase, Maryland. The centre has had a lot of press and publicity, so there is a lot of hype about the place — which helped with recruitment. Now we need to start publishing results to show the public what we are doing. ■

INTERVIEW BY VIRGINIA GEWIN